

ESSAYS ON EDUCATIONAL INEQUALITIES

by

Hojung Lee

A dissertation proposal submitted to the Faculty of the University of Delaware
in partial fulfillment of the requirements for the degree of Doctor of Philosophy in
Education & Social Policy.

Spring 2025

© 2025 Hojung Lee
All Rights Reserved

ESSAYS ON EDUCATIONAL INEQUALITIES

by

Hojung Lee

Approved: _____
Laura Desinome, Ph.D.
Director of Research for CEHD

Approved: _____
Rena Hallam, Ph.D.
Interim Dean of School of Education

Approved: _____
Louis F. Rossi, Ph.D.
Vice Provost for Graduate and Professional Education and
Dean of the Graduate College

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Kenneth A. Shores, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Florence Xiaotao Ran, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Adrienne Lucas, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Nora Gordon, Ph.D.
Member of dissertation committee

ACKNOWLEDGEMENTS

This work would not have been possible without the help of numerous people in my academic and personal community, some of whom I particularly acknowledge here. First and foremost, I express my deepest gratitude to my advisor, Dr. Kenneth Shores, who has fundamentally shaped who I am as a scholar and researcher. As his first advisee, I have been fortunate to experience his truly exceptional mentorship. While his feedback can be direct and unvarnished, this honesty stems from his genuine commitment to his students' professional growth and success. Throughout my five years at UD, he has consistently pushed me to think more deeply, write more clearly, and approach research with both methodological rigor and real-world relevance. More than just technical skills, he has taught me what it means to be a principled researcher who conducts honest, meaningful scholarship. His mentorship transcends typical academic guidance – he has shown me how to be not just a better scholar, but a better thinker and a more conscientious person. The advisor-advisee relationship we have built, grounded in academic respect and trust, has been one of the most transformative experiences of my academic journey.

My deep appreciation extends to the rest of my dissertation committee. I am particularly grateful to Dr. Florence Ran, who has been an invaluable mentor throughout my Ph.D. journey. Her academic guidance and emotional support, coupled with her warm personality, helped me navigate the challenges of this journey. I am also indebted to Dr. Adrienne Lucas, who provided great feedback on my dissertation despite her significant responsibilities as Chair of the Department of

Economics. Chapter 1 particularly benefited from the expertise of Dr. Nora Gordon, a leading scholar in U.S. Title I research, who served as my external committee member and whose work has profoundly influenced my research direction.

I also deeply benefitted from the wonderful community and faculty in the College of Education and Human Development and Joseph R. Biden Junior School of Public Policy and Administration. The program of Education and Social Policy came to exist through the dedication of Drs. Laura Desimone and Sarah Bruch. I also appreciate the support from Christina Johnston, Alexis Milam, and Kristin Hatfield, who are irreplaceable but often not appreciated as they should be.

I owe a profound debt of gratitude to my dearest friends, who became my family when I needed it most. Yeonsin Kwon, who has been a constant in my life for fifteen years, has supported me through every challenge. I am not merely proud to call her my best friend; I owe my life's trajectory to her unwavering support. Similarly, Changhee Lee and Hoon Jeong have been more than friends and mentors – their faith in me, demonstrated through both emotional and practical support, made this journey possible.

Lastly, I thank my family members for their support throughout this journey. I am also grateful to Kyungdong Kim, whose consistent support and extraordinary patience have been invaluable to me both personally and professionally. While space constraints prevent me from naming everyone who has contributed to this journey, I am profoundly grateful for all the love, friendship and encouragement that have shaped both this work and my life.

TABLE OF CONTENTS

LIST OF TABLES	x
LIST OF FIGURES	xiii
ABSTRACT	xiv
 Chapter	
1 INTRODUCTION TO THE DISSERTATION	1
1.1 Paper One: Supplanting or Supplementing: The Stickiness of Title I Revenues in Post-Adequacy Era	2
1.2 Paper Two: Does Corequisite Remediation Work for Everyone? An Exploration of Heterogeneous Effects and Mechanisms (with Dr. Florence Xiaotao Ran)	3
1.3 Paper three: The Architecture of Expected Wage Gaps: Between- and Within-School Sources of Career Education Inequality (With Dr. Kenneth A. Shores and Arielle Lentz)	4
1.4 Contribution	4
 2 PAPER ONE: SUPPLANTING OR SUPPLEMENTING: THE STICKINESS OF TITLE I REVENUES IN POST-ADEQUACY ERA	6
2.1 Introduction	7
2.2 Literature Review	13
2.3 Background on Title I	16
2.3.1 Evolution of Title I Funding Structure	17
2.3.2 Calculation vs. Distribution of Title I Funds	17
2.3.3 Title I Formula and Key Eligibility Thresholds	18

2.3.4	Changes in Poverty Measurement and Implications	18
2.4	Data	19
2.4.1	Title I Allocation Information	19
2.4.2	District Finances, Enrollment, and Demographic Information	21
2.4.3	Tax and Bond Election Data	23
2.5	Empirical Strategy	24
2.5.1	Endogeneity challenges in Title I analysis	24
2.5.2	Identification Challenges in Multi-Threshold Title I Formula	25
2.5.3	Stacked Regression Discontinuity Approach	26
2.5.4	Bandwidth Selection Strategy	27
2.5.5	Estimation	28
2.5.6	Validating the Stacked RD Design	30
2.6	Results	32
2.6.1	First Stage Results: Change in Title I Allocation	32
2.6.2	Effects on Districts' Finance	34
2.6.3	Effects on Local Tax and Bond Elections	35
2.6.4	Implication on State Funding Policies	37
2.7	Discussion and Conclusion	40
3	PAPER TWO: DOES COREQUISITE REMEDIATION WORK FOR EVERYONE? AN EXPLORATION OF HETEROGENEOUS EFFECTS AND MECHANISMS (WITH DR. FLORENCE XIAOTAO RAN)	55
3.1	Introduction	56
3.2	Literature Review and Conceptual Framework	60
3.2.1	Previous literature on developmental education	60
3.2.2	Conceptual framework of the effects of corequisite DE	61
3.3	Context & Data	64
3.3.1	State & institutional context	64

3.3.2	Data Description	66
3.4	Empirical Strategy	68
3.4.1	Identification Strategy	68
3.4.2	Validity of DID estimates	70
3.5	Results	73
3.5.1	Effects of corequisite remediation	73
3.5.2	Implications of corequisite reform effects	77
3.5.3	Exploration of effects mechanism	78
3.6	Conclusion & Discussion	81
4	PAPER THREE: THE ARCHITECTURE OF EXPECTED WAGE GAPS: BETWEEN- AND WITHIN-SCHOOL SOURCES OF CAREER EDUCATION INEQUALITY (WITH DR. KENNETH A. SHORES AND ARIELLE LENTZ)	92
4.1	Introduction	93
4.2	Research Context: Career Technical Education in Delaware	96
4.3	Conceptual Framework: Expected Wages as a Measure of Educational Stratification	97
4.4	Data	99
4.5	Methods	101
4.6	Results	105
4.6.1	Research Question One: Magnitude and Distribution of Inequality	105
4.6.2	Research Question Two: Within or Between School Factors	106
4.6.3	Research Question Three: Predictors of School-Level Gaps in Wage Inequality	110
4.7	Policy Applications	111
4.8	Conclusion	113
 Appendix		
A	PAPER ONE APPENDIX TABLES AND FIGURES	135
B	PAPER TWO APPENDIX TABLES AND FIGURES	142

C PAPER THREE TECHNICAL APPENDIX: MEDEL SENSITIVITY CHECK	156
D PAPER THREE APPENDIX TABLES AND FIGURES	160

LIST OF TABLES

2.1	Title I Grant Eligibility and Weighting Formula	45
2.2	District Fiscal and Nonfiscal Characteristics (Summary Statistics)	46
2.3	District Fiscal and Nonfiscal Characteristics (Stacked Sample) . .	47
2.4	McCrary Density Test	48
2.5	Covariate Balance Test	49
2.6	First Stage: Effects of Title I Formula Cutoffs on Title I Revenues	50
2.7	Effects on District Finance (Per Pupil)	51
2.8	Effects on Tax and Bond Elections	52
2.9	Effects on Tax and Bond Elections (By Title I Quartile)	53
2.10	Association Between State Funding Progressivity and Title I Effects on District Finances	54
3.1	Summary Statistics (2010–11 to 2018–19 Cohorts)	84
3.2	Effects of corequisite remediation on first-year gateway course completion	85
3.3	Effects of corequisite remediation on first-year gateway course completion by placement test scores	86
3.4	Effects of corequisite remediation on subsequent college-level course outcomes (by end of Y2)	87

3.5	Effects of corequisite remediation on credit accumulation, persistence, transfer and completion (by end of Y3)	88
3.6	Timing of first gateway course enrollment before and after corequisite reform by test score group	89
3.7	Number of developmental credits enrolled and earned before and after corequisite reform by test score group	90
3.8	Association between proportion of on-level peers and first-year gateway course completion rates	91
4.1	Mean expected wages and student enrollment by career cluster . .	116
4.2	Regression results from multilevel mixed-effects models	117
4.3	Regression results from multilevel mixed-effects model	118
4.4	Decomposing Expected Wage Gap among Concentrators (w/ Test Scores)	119
4.5	Multilevel Mixed Effects Model with School-Level Characteristics	120
A.1	McCrary Density Test (Count Sample)	138
A.2	First Stage: Effects of Title I Formula Cutoffs on Title I Revenues	139
A.3	Effects on District Finance (Per Pupil, PML)	140
A.4	Effects on District Finance, Federal Source (Per Pupil)	141
B.1	Evidence on the Effects of Traditional and Corequisite Developmental Education	144
B.2	Summary statistics of different samples	150
B.3	Covariate balance test	152
B.4	Effects of corequisite remediation on first-year gateway course completion: Robustness check (excluding data from academic year 2019–20)	153

B.5	Effects of corequisite remediation on credit accumulation, persistence, transfer, and completion by end of Y3: Robustness check (excluding data from academic year 2019–20)	154
B.6	Effects of corequisite remediation on gateway course completion and downstream outcomes: Robustness check (excluding students from high schools with SAILS program)	155
D.1	Regression results from fixed-effects model	161
D.2	Correlation & SD of Anticipated Wage Gaps (Weighted Coefficients)	162
D.3	Multilevel Mixed Effects Model with School-Level Characteristics	163
D.4	Multilevel Mixed Effects Model with School-Level Characteristics	164

LIST OF FIGURES

2.1	Distribution of Running Variables	43
2.2	Density of observations (proportion criteria)	44
3.1	Proportions of on-level peers in gateway course sections by academic year	83
4.1	Gender wage gap distribution and female concentration in CTE pathways.	115
A.1	Density of observations (count criteria)	136
A.2	Density of Proportion of Eligible Kids, by State	137
B.1	Event study estimators comparison — effects on first-year gateway completion.	143

ABSTRACT

This dissertation presents three empirical studies examining critical issues in U.S. education policy and finance—specifically, how policy design and institutional contexts either mitigate or reproduce inequities in school funding, instructional quality, and career readiness. Each chapter addresses a distinct dimension of K–16 schooling in the United States, and collectively they offer insights into the multifaceted ways that resources, instructional reforms, and career preparation intersect with equity goals.

Paper One investigates Title I, the largest federal educational grant aimed at supporting low-income students in K–12 schooling. Using a stacked regression discontinuity design, I examine whether Title I truly supplements district budgets or crowds out state and local revenue. The findings suggest that, while large-scale displacement of operating funds is not apparent, some districts respond by reducing capital outlays and bond elections—especially those with larger Title I allocations. These results underscore how evolving institutional characteristics can affect district-level fiscal behavior.

Paper Two (with Dr. Florence Ran) analyzes corequisite remediation in community colleges, a widely adopted reform intended to improve the success of academically underprepared students in college-level coursework. Using administrative data from the Tennessee community college system and employing difference-in-differences and event study analyses, this study shows that corequisite models improve gateway course completion for students across a range of academic readiness levels. However, these positive impacts vary by student preparedness, and

students were more likely to drop out of the public college system after the reform—highlighting the need for additional supports.

Paper Three (with Dr. Kenneth Shores and Arielle Lentz) examines high school Career and Technical Education (CTE) programs and their potential role in shaping future wage inequality. Linking Delaware’s administrative data with occupational wage information, this paper explores wage disparities across student subgroups and investigates the sources of these gaps. The results reveal notable differences in expected wages: female–male wage gaps arise largely from within-school factors, whereas racial and socioeconomic gaps primarily reflect across-school differences. These patterns suggest that high school career programs, though seemingly neutral, may inadvertently produce gaps in students’ future earnings.

Taken together, these three papers illustrate how equity considerations are woven into the fiscal, curricular, and structural facets of U.S. education. By identifying unintended consequences of policy interventions, this dissertation provides policy-relevant evidence on how educational systems can better serve all students.

Chapter 1

INTRODUCTION TO THE DISSERTATION

Education policy in the United States continues to confront entrenched forms of inequality. In the 2020–21 school year alone, public K–12 schools spent \$927 billion, while public postsecondary institutions expended another \$450 billion (National Center for Education Statistics, 2023, 2024), but significant gaps persist across different regions and student subgroups in funding, instructional quality, and career readiness—ultimately translating into unequal long-term economic outcomes. A central question arises: do these large-scale policy efforts advance the equity goals they were designed to achieve, or do they at times sustain or even create new forms of inequality?

This dissertation explores these questions by examining three distinct yet interconnected policy arenas—federal Title I funding, developmental education reform, and high school career and technical education (CTE). Although each policy area targets different stages in the K–16 pipeline, they share an overarching concern: how does institutional context shape whether a policy narrows or exacerbates existing disparities? For instance, federal grants intended for low-income school districts may be absorbed into local budgets in unanticipated ways, remedial reforms at the college level may benefit some student groups more than others, and CTE expansions may inadvertently direct certain populations into lower-wage career paths.

Prior research on educational interventions often focused on average treatment effects—measuring overall program impacts without fully accounting for potential heterogeneity (Angrist et al., 2002; Krueger, 2003). Recent scholarship, however, underscores that policy outcomes are deeply contingent upon local conditions, resource constraints, and student demographic characteristics (Candelaria & Shores, 2019; Dynarski et al., 2018; Jackson et al., 2016). Building on this recognition, I situate my dissertation at the crossroads of three literatures: (1) fiscal federalism and intergovernmental grants in education (Gordon, 2004; Inman, 2008), (2) mechanisms and structural reforms in developmental education (Kane et al., 2019; Scott-Clayton & Rodriguez, 2015), and (3) the measurement of educational stratification (Domina et al., 2017; Reardon et al., 2019, 2014). Across these frameworks, a recurring theme emerges—policy can produce varied and sometimes unexpected outcomes across different contexts and populations.

By applying these theoretical lenses, this dissertation presents three empirical studies that collectively illustrate how resource allocation, instructional practices, and career preparation can either mitigate or perpetuate educational inequities. Each chapter addresses a distinct dimension of U.S. schooling, from early K–12 finance to postsecondary readiness and high school career pathways. Taken together, the chapters highlight the importance of looking beyond aggregate outcomes: to understand whether policies truly expand opportunity, we must examine how local institutions, specific student groups, and policy design factors interact in practice.

1.1 Paper One: Supplanting or Supplementing: The Stickiness of Title I Revenues in Post-Adequacy Era

Paper one investigates Title I, the largest federal educational grant that aims to support low-income students in K-12 schooling. Using a stacked regression discontinuity design that leverages multiple discontinuities in the Title I formula, I examine whether Title I truly supplements district budgets or crowds out state and

local revenue in the "post-adequacy" era of school finance. While earlier research documented substantial local fiscal substitution in response to Title I (Cascio et al., 2013; Gordon, 2004), my findings suggest that the institutional context of modern state funding systems has altered this relationship. Large-scale displacement of operating funds is no longer apparent, but a more subtle form of fiscal adjustment emerges: districts respond by reducing capital outlays and bond elections, especially districts with larger Title I allocations. This paper contributes to the fiscal federalism literature by demonstrating how changing institutional characteristics affect intergovernmental grant dynamics and highlights the importance of examining multiple dimensions of fiscal response beyond basic revenue substitution.

1.2 Paper Two: Does Corequisite Remediation Work for Everyone? An Exploration of Heterogeneous Effects and Mechanisms (with Dr. Florence Xiaotao Ran)

Paper two analyzes the corequisite remediation in community colleges, a widely adapted reform in community colleges to improve student success in college-level coursework for academically underprepared students. Using administrative data from the Tennessee community college system and employing difference-in-differences and event study designs, this study shows that corequisite models improve gateway course completion across students from various academic preparedness levels. However, the positive impacts differ by student readiness, with stronger effects for higher-scoring students in math and lower-scoring students in English. Furthermore, students were more likely to drop out of the public college system after the corequisite reform, particularly those with the lowest test scores. These findings advance our understanding of how structural reforms in developmental education produce heterogeneous effects and suggest the need for additional support targeted to the most academically vulnerable students.

1.3 Paper three: The Architecture of Expected Wage Gaps: Between- and Within-School Sources of Career Education Inequality (With Dr. Kenneth A. Shores and Arielle Lentz)

Paper three examines high school career technical education (CTE) program and its potential role in shaping future wage inequality. Using Delaware’s administrative data linked with occupational wage information, this paper explores potential wage inequalities between student subgroups and identifies their sources through decomposition analyses. Results show notable gaps in expected wages by student demographics, with gender wage gaps primarily stemming from within-school factors, while gaps between different racial groups and between low-income and non-low-income students originate primarily from between-school factors. By separating within-school from between-school mechanisms of inequality, this paper provides insights into how seemingly neutral career preparation programs may produce unintentional wage gaps through both programmatic design and resource allocation patterns across schools.

1.4 Contribution

The connections among these three papers extend beyond their shared methodological approaches and focus on equity. Each paper examines a different point in the educational pipeline—K-12 funding, postsecondary instruction, and career preparation—yet together they reveal how institutional structures mediate policy effects across educational contexts. For example, the findings on Title I’s effects on capital investments connect to the CTE paper’s emphasis on resource distribution across schools, as both demonstrate how financial decisions shape educational opportunities in ways that can reproduce inequality. Similarly, the corequisite remediation paper’s findings on differential effects by academic preparation levels parallel the CTE paper’s results on how program selection varies by student demographics,

highlighting how educational structures can create divergent pathways for different student groups.

Taken together, these three papers demonstrate how equity considerations are woven into fiscal, curricular, and structural facets of the U.S. education system. They collectively suggest that even well-intentioned policies designed to improve educational opportunities may have unintended consequences that reinforce existing inequalities if not carefully designed with attention to institutional contexts and potential heterogeneous effects. By identifying these mechanisms, this dissertation provides policy-relevant evidence for creating more effective and equitable educational interventions. The findings underscore that if equity is the guiding principle, policymakers and educators must remain vigilant to the ways in which budgets, curricula, and career tracks can become engines for both inclusion and inequality.

In this sense, the dissertation as a whole argues for a deeper, more nuanced appreciation of “equitable policy success.” Rather than merely taking aggregate improvements as confirmation that a program “works,” researchers and practitioners need to investigate which students or districts are truly benefiting and which ones might be left behind or unintendedly disadvantaged. By linking fiscal, instructional, and career contexts under a single lens, these studies ultimately emphasize that reducing educational inequality requires a multi-dimensional approach, one that carefully attends to how each layer of the system shapes the distribution of opportunity.

Chapter 2

PAPER ONE: SUPPLANTING OR SUPPLEMENTING: THE STICKINESS OF TITLE I REVENUES IN POST-ADEQUACY ERA

Abstract This paper examines how school districts adjust their financial behavior in response to federal Title I funding. Using a stacked regression discontinuity design that leverages multiple discontinuities in the Title I formula, this study estimates the effects of Title I eligibility on district revenues and expenditures, providing insights into fiscal federalism. The findings indicate no substantial evidence of crowding out local or state funding efforts in response to Title I allocations, with local revenue increasing by \$182 per pupil (95% CI: -\$70 to \$434) and state revenue by \$76 per pupil (95% CI: -\$48 to \$200). These confidence intervals rule out substantial displacement effects, suggesting that recent institutional changes have successfully prevented the substitution of federal educational grants. However, the analysis reveals that Title I funding is associated with a decrease in capital expenditures of \$83 per pupil ($p < 0.10$) and a 2.8 percentage point reduction in bond elections, particularly in districts receiving higher levels of Title I funding. These results show how state fiscal policies and federal grants work together to shape local government behavior, with fiscal adjustments appearing in capital investments rather than operating budgets.

2.1 Introduction

Educational funding plays a critical role in providing equal opportunities, yet the distribution of resources has been historically unequal (Card & Payne, 2002; Jackson et al., 2016; Lafortune et al., 2018; Reeves, 2013). The cornerstone of federal K-12 aid is Title I of the 1965 Elementary and Secondary Education Act, created as part of the War on Poverty. Title I is the largest federal program supporting elementary and secondary education. In the 2023 fiscal year, the total discretionary budget for the U.S. Department of Education was \$79.6 billion, with Title I alone accounting for \$19 billion – more than 23% of the total budget. The fundamental goal of Title I is to supplement the resources available to disadvantaged students, ensuring additional services and support beyond what state and local funds provide. Indeed, Shores et al. (2021) show that Title I can fill the gap between low- and non-low-income students gap if well-targeted.

However, a persistent concern is whether Title I dollars truly add to local school budgets or merely supplant funds that state and local authorities would otherwise spend. If federal aid simply replaces local funding rather than adding to it, then the intended beneficiaries—disadvantaged students—receive no additional resources, and the program fails to achieve its primary purpose. The federal government has recognized this concern by implementing “supplement, not supplant” provisions that require Title I funds to add to, rather than replace, local funding. Maintenance-of-effort provisions further require districts to maintain at least 90% of their prior year education spending from state and local sources (New America, n.d.). Despite these requirements, research has shown that Title I revenues have been crowded out by state and local effort.

Specifically, studies examining Title I’s early implementation and its effects through the 1990s documented that federal dollars largely replaced rather than added

to local education spending. Cascio et al. (2013) show that each Title I dollar increased district spending by \$0.50, but decreased local revenue by about \$0.33. Similarly, Gordon (2004) shows districts offset Title I grants almost entirely within three years, estimated by \$1.2 crowd-out per Title I dollar per pupil. These papers raise major concerns about whether Title I was achieving its statutory purpose of providing additional resources to disadvantaged students. However, these studies took place in an institutional era that predates the major transformation of education finance systems that began in the late 1990s. The contemporary landscape in which Title I operates has been fundamentally changed by adequacy-focused reforms, stronger federal accountability frameworks, and changes in poverty measurement methodologies, creating a compelling case to reassess Title I's effectiveness in the current environment.

Importantly, much of the existing evidence on Title I's effects comes from periods before the late 1990s and early 2000s, which marked the beginning of significant changes in school finance systems across the United States. The so-called "adequacy era," which scholars identify as beginning with the *Rose v. Council for Better Education Case* in 1989, was largely driven by a wave of state court cases, court rulings, and legislative reforms that rejected the notion that equal funding was sufficient to provide equal educational opportunity. Instead, the courts found that the constitutional basis for education funding required a guarantee that all students receive resources adequate to achieve state-established educational standards. In response, states implemented new funding formulas that typically guaranteed each district a base funding level considered "adequate", while often providing extra support for low-income students and English language learners (Darling-Hammond, 2019; Lafortune et al., 2018). These policies aimed to direct more resources toward less affluent districts and, in some cases, equalized funding across districts by ensuring that standard tax rates would generate similar funding regardless of local property value (Hoxby, 2001).

In addition to the progressivity of state revenues mechanically incorporated into most state funding formula, several additional institutional features of the adequacy era may have altered the dynamics of how districts respond to Title I funding. First, adequacy-based funding formulas often incorporated stronger state oversight of district finances, including more explicit maintenance-of-effort provisions that limited districts' ability to reduce their own funding contributions. Second, many states imposed property tax floors and ceilings during this period, constraining the ability of local districts to adjust tax rates in response to changes in federal aid (Verstegen, 2004; Verstegen & Jordan, 2009). Third, the No Child Left Behind Act of 2001 introduced a new accountability framework for Title I, requiring that funds be tied to specific school improvement strategies and measurable student outcomes. This heightened accountability could further discourage the diversion of Title I dollars from their intended educational purposes.

In 1999, another important change occurred to the Title I program itself, when Title I transitioned from the decennial US Census to the annual Small Area Income and Poverty Estimates (SAIPE) data for determining district-level eligibility counts. Previously, Title I calculations relied on decennial census data as the primary poverty measure, creating abrupt funding shocks during census updates but allowing districts time to adjust their funding efforts in subsequent years between the census cycles. With the transition to annual SAIPE data, districts now face more frequent adjustments, potentially reducing their ability to adapt funding in response to Title I changes and limiting districts' fiscal flexibility and ability to predict future allocations accurately.

Given these substantial institutional changes to state school funding policy, federal No Child Left Behind policy, and Title I itself, it is unclear whether the relationship between Title I funding and local fiscal behavior observed in earlier decades still holds in the current policy environment. This paper contributes to our understanding of this relationship by revisiting the impact of Title I on school

funding in the post-adequacy era. I examine district-level data from 2008 through 2017, when adequacy-based formulas and strengthened accountability measures were firmly established across most states.

To overcome the empirical challenges inherent in studying the causal effects of Title I, I employ a novel stacked regression discontinuity design that leverages multiple discontinuities in the Title I funding formula. Title I's complex formula creates several distinct eligibility thresholds and multiple weighted count thresholds for allocation purposes. Rather than conducting separate analyses at each threshold — which would produce isolated estimates with limited statistical power — the stacked approach treats each threshold as a separate quasi-experiment and combines them into a unified framework. This methodology has several advantages over traditional RD approaches: it maintains clean control groups by preventing contamination between thresholds, increases statistical power by pooling information across multiple discontinuities, and allows for the estimation of an overall treatment effect while accounting for threshold-specific heterogeneity. The stacked RD approach thus provides more robust and generalizable inferences about Title I's effects than previous single-threshold studies.

Complementing this methodological approach, this study also benefits from administrative data that provides precise counts of Title I eligible students exactly as used in the official funding calculations. This data precision is crucial for regression discontinuity analysis, as it eliminates measurement error that could otherwise compromise identification. My empirical analysis first establishes that at these formula thresholds, Title I allocation increases by \$64.68 per eligible child (approximately 4.7% increase from the baseline mean), creating a meaningful discontinuity for causal inference. With this variation, I then examine how districts adjust their fiscal behavior in response to these additional federal resources. Additionally, I disaggregate elementary and secondary expenditure into instruction and capital outlays to examine whether Title I funding influences specific spending patterns. Instruction is

central to Title I’s goal and accountability requirements under No Child Left Behind, which emphasized student achievement in test subjects. Capital outlays are interesting because they are often financed through separate mechanisms like bond elections, rather than through general tax elections, making them more discretionary and potentially more subtle to fiscal substitution.

My findings reveal a fundamental shift in how local governments respond to Title I funding in the post-adequacy era. In contrast to earlier studies, I find no evidence that Title I crowds out state and local education spending. Specifically, Title I eligibility increases state revenue and local revenue by \$75 and \$182 per pupil, respectively, though these effects are not statistically significant. This absence of crowd-out stands in stark contrast to previous research by Gordon (2004) and Cascio et al. (2013), which documented substantial substitution effects, and suggests that the institutional constraints of the adequacy era have successfully prevented the direct substitution of federal educational grants observed in earlier decades.

However, my analysis also uncovers evidence of a subtle form of fiscal substitution at the district level. While traditional measures of crowd-out, such as reductions in state and local education revenue, do not appear, districts at the Title I margin reduce capital expenditures by \$83 per pupil (approximately 6.4% of mean capital outlays, $p < 0.1$). This decline is driven in part by lower bond referendum activity, particularly among districts receiving higher levels of Title I funding. Districts above Title I eligibility thresholds are 2.8 percentage points less likely to hold bond elections (6.9% decrease from the baseline probability), with this effect concentrated among districts in the highest quartile of Title I funding per eligible child, which show a 4.7 percentage point reduction in bond elections (13.9% decrease from baseline).

These findings align with tax salience theory (Brunner et al., 2018, 2021; Cabral & Hoxby, 2012), which suggests that when regulatory constraints limit operational budget flexibility, policy responses emerge in less visible tax channels.

The observed pattern – where Title I affects lower-salience bond referenda but not higher-salience tax elections — indicates that districts are still responding to federal aid by seeking tax relief, but through less direct mechanisms that do not violate maintenance-of-effort provisions or other operational funding requirements.

Institutional differences in how states finance public education may also change district-level responses to federal aid. To test this relationship, I construct a measure of funding progressivity that captures how much more dollars are allocated to Title I-eligible kids compared to non-eligible kids. This progressivity measure serves as a proxy for how strongly a state’s funding system already prioritizes disadvantaged students, potentially influencing how districts respond to additional federal resources targeted at the same population. I also include indicators for state property tax floor and ceiling provisions to account for additional institutional constraints on local fiscal response. I then test whether the effect of Title I on district financial behavior varies systematically across these state policy environments.

I find that in states with higher elementary/secondary expenditure progressivity, the negative effects of Title I on capital outlays are significantly stronger (-0.157 , $p < 0.05$), indicating that for each additional dollar allocated to Title I-eligible students over non-eligible students in a state’s funding system, districts reduce their capital expenditures by approximately 16 cents when receiving Title I funds. In contrast, expenditure progressivity shows no significant relationship with how Title I affects state or local revenue (coefficients of -0.010 and 0.050 respectively). This suggests that regardless of how progressive a state’s spending is toward disadvantaged students, Title I has a similar (non-displacing) effect on operational funding sources across different state systems, though additional aid for Title I eligible children reduces demand for capital outlays. Likewise, neither property tax floor nor ceiling provisions systematically moderate district revenue responses to Title I. Taken together, these findings suggest that while adequacy-era reforms have strengthened

baseline operational support for disadvantaged districts, they have created a consistent pattern where districts receiving Title I funds reduce capital investments rather than operational spending, regardless of specific state revenue policy structures.

The remainder of this paper proceeds as follows. Section 2 reviews the literature on intergovernmental grants and school finance in greater detail. Section 3 provides background on Title I and its evolution. Section 4 presents the data and identification strategy. Section 5 presents the findings, and Section 6 discusses their implications and concludes.

2.2 Literature Review

A large body of research has examined whether intergovernmental grants like Title I lead to higher total spending on education or whether recipient governments offset the grant by reducing their own contribution. Early theoretical work on fiscal federalism predicted a “flypaper effect” in which grant money sticks where it hits – implying that dollar of federal aid should boost public spending by something close to a dollar (Hines Jr & Thaler, 1995; Inman, 2008; Oates, 1999). However, if local officials treat federal dollars as a substitute for their own funds, the increase in spending could be far less. For instance, when school districts (or other local governing bodies) receive new federal grants, they might lower local tax effort or a state might redirect existing state aid, thereby crowding out the intended additional spending (Baicker & Gordon, 2004; Chakrabarti & Setren, 2011; Fisher & Papke, 2000; Nguyen-Hoang & Yinger, 2020; Steinberg et al., 2016).

Empirical findings on Title I have been mixed. Several studies in earlier decades found evidence of substantial crowd-out of Title I funds, suggesting that federal aid often freed up state/local money for other uses instead of increasing school spending dollar-for-dollar. For example, Gordon (2004) finds that in the 1990s, districts offset federal education grants almost entirely within a few years, such that after about three years there was no net increase in school spending from Title

I. Similarly, Van der Klaauw (2008), studying New York City schools, reports no discernible effects of Title I on per-pupil expenditure, consistent with local budgets adjusting to absorb the federal aid. In more recent analysis, Matsudaira et al. (2012) also find no robust association between Title I eligibility and per-pupil spending, reinforcing the view that Title I funds have often been fully supplanted by reductions in other funding. On the other hand, in the context of Title I's initial introduction, Cascio et al. (2013) show that Southern school districts increased spending by about \$0.50 for each federal dollar received.

Thus, prior literature indicates that Title I sometimes supplements and sometimes supplants existing funding – with the degree of crowd-out varying by context and time period. A key insight is that institutional features and incentives at the state/local level can mediate the effect of federal grants. For instance, maintenance-of-effort requirements and political or community pressures can encourage districts to use federal funds for new programs, whereas budgetary fungibility and lax enforcement can enable substitution.

Given these varied findings in the Title I literature, it is important to consider how broader structural changes in education finance might alter the way districts respond to federal funding. The institutional environment for education funding has undergone dramatic transformation since many of the classic Title I studies were conducted. These reforms were aimed at ensuring that all districts, especially those serving low-income communities, have sufficient resources to provide an adequate education. In contrast to earlier equity-focused reforms, which often sought to equalize spending across districts, the adequacy-driven changes explicitly targeted funding adequacy and typically resulted in large increases in state aid to previously under-resourced districts. Research finds that these post-1990 school finance reforms led to sharp, immediate, and sustained increases in per-pupil spending in low-income districts (Candelaria & Shores, 2019; Lafortune et al., 2018). One mechanism for accomplishing this allocation is by a foundation grant system, which guarantees

each district a base funding level (foundation) deemed adequate and fills the gap between that amount and what the district can raise with a standard local tax effort (Darling-Hammond, 2019; Lafortune et al., 2018). As a consequence, the state's share of education funding rose, and local spending disparities were reduced.

These reforms often come with additional institutional constraints intended to prevent the diversion of the new funds, for example, by requiring local districts to maintain or increase their own funding efforts and to spend the funds on specified educational inputs (Verstegen & Jordan, 2009). In effect, the combination of guaranteed funding levels, substantial state aid, and mandated local effort created a structure in which extra resources could not be easily directed away and, thus, any additional dollar was more likely to result in higher spending on education (Card & Payne, 2002; Duncombe et al., 2003; Hoxby, 2001). At the same time, federal education policy was evolving: the No Child Left Behind Act of 2001 reauthorized and expanded Title I, introducing stricter accountability for student outcomes and requiring that Title I funds be tied to school improvement strategies (Domanico, 2024). In principle, this heightened accountability could further discourage the misuse of Title I dollars, pressuring districts to devote the funds to genuine instructional enhancements rather than budget relief.

Despite the substantial changes in education finance during the post-adequacy era, relatively little research has directly examined how these institutional reforms might have altered the relationship between Title I funds and local fiscal behavior. The literature on school finance reforms has documented increased spending in low-income districts (Candelaria & Shores, 2019; Lafortune et al., 2018), but hasn't specifically addressed how these reforms might interact with federal Title I funding. Similarly, studies of Title I effectiveness (Cascio et al., 2013; Gordon, 2004) have not been updated to reflect the new institutional environment created by adequacy-focused reforms. This creates an important empirical question: whether

the effectiveness of Title I in the post-adequacy era differs fundamentally from patterns documented in earlier periods.

An important strand of literature that may help explain varied responses to intergovernmental grants examines tax salience and fiscal visibility. Cabral and Hoxby (2012) demonstrate that property tax rates decrease when taxes are more salient to residents, suggesting local governments respond differently to highly visible versus less visible fiscal adjustments. Similarly, Brunner et al. (2018, 2021) show that voter support for educational spending measures depends significantly on how tax implications are framed and presented. This research suggests that even when direct budget substitution is constrained by institutional rules, local governments might respond to intergovernmental grants through less visible fiscal channels. Baicker and Gordon (2004) provide evidence for this possibility, showing that state education finance reforms influenced not only direct educational spending but also other categories of local government expenditure. Together, these studies highlight the importance of examining multiple fiscal response channels when assessing the impact of federal aid programs like Title I.

2.3 Background on Title I

Title I, established under the Elementary and Secondary Education Act (ESEA) of 1965, serves as the federal government’s primary mechanism for promoting educational equity by supporting programs tailored to meet the needs of educationally deprived children from low-income families. As part of President Johnson’s War on Poverty, the ESEA aimed to tackle poverty at its root by expanding economic opportunities for disadvantaged populations through education. Title I marked a significant milestone in U.S. education finance by shifting the focus of educational inequality debates from racial disparities to economic ones, as reflected in its allocation process.

2.3.1 Evolution of Title I Funding Structure

The current Title I allocation formula was finalized under the No Child Left Behind Act (NCLB) of 2002, which reauthorized the ESEA. Initially, the Title I funding amount was determined through a straightforward formula that multiplied the state per pupil expenditure (SPPE) by the number of children from low-income families. Since Title I is intended to be a supplemental program, districts receive 40 cents per SPPE dollar. The original ESEA defined Title I eligibility for children aged 5-17 from families earning less than \$2,000 annually or receiving Aid to Families with Dependent Children (AFDC). Over time, the eligibility criteria expanded to include children who are neglected, delinquent, or in foster care through the Elementary and Secondary Education Amendment of 1966 (Skinner & Riddle, 2020).

2.3.2 Calculation vs. Distribution of Title I Funds

A critical aspect of Title I funding is that while district-level characteristics determine eligibility and funding amounts, the actual distribution follows a two-step process. First, the U.S. Department of Education calculates district-level allocations using allocation formulas that account for the number of low-income students and state-level per-pupil expenditures. However, rather than directly disbursing these funds to districts, the federal government channels them through state education agencies, which then distribute the funds to local school districts following additional state-level allocation rules.

This distinction between calculation and distribution has important implications for empirical analysis. At the first stage, states receive different per-eligible-pupil amounts due to variations in state per-pupil expenditures, hold-harmless provisions, and adjustments for state education finance policies. This creates disparities in per-pupil Title I funding across states, even for districts with similar poverty rates. At the second stage, within-state allocation rules determine how districts

receive their share of the state’s total allocation. Understanding this two-step allocation process provides context for how Title I funding reaches districts and reveals potential sources of variation in funding distribution.

2.3.3 Title I Formula and Key Eligibility Thresholds

The current allocation of Title I is based on the formula established in 2002, which includes a complex system of four different grants, each with its own allocation formula and eligibility criteria. Basic Grants, which typically account for 45% of the total Title I budget, are allocated to districts with at least ten “formula children” and where this number exceeds 2% of the total 5-17 population. Concentration Grants, representing approximately 10% of the total budget, are distributed to districts with over 6,500 formula children or where this number exceeds 15% of the total school-age population. Targeted Grants and Education Finance Incentive Grants (EFIG), each accounting for about 25% of the budget, are awarded to districts with at least ten formula children and where this number exceeds 5% of the 5-17 population. Additionally, the weighted calculation of formula children adds further complexity to the allocation of Targeted Grants and EFIG. Details of Title I grant eligibility and weighting criteria are documented in Table ??.

[Table ?? about here]

This complex formula structure, with its multiple cutoffs and weighting mechanisms, inherently generates exogenous variation in funding allocation. These discontinuities form the basis of the stacked regression discontinuity design (RDD) employed in this study, as they create natural experiments around each eligibility threshold.

2.3.4 Changes in Poverty Measurement and Implications

Beginning from the 1999 income year, a significant shift in Title I implementation occurred when the annual Small Area Income and Poverty Estimates (SAIPE)

data collection was used to determine district-level Title I eligibility counts. Prior to this, Title I calculations relied on decennial census data as the primary poverty measure. The use of a poverty measure updated once every ten years created abrupt funding shocks during census updates but allowed districts time to adjust their funding efforts in subsequent years. Gordon (2004) uses this mechanism and finds that local governments substantially crowd out changes in Title I after three years of poverty measure change.

However, with the transition to annual SAIPE data, districts now face more frequent adjustments, reducing their ability to adapt funding in response to Title I changes. This shift could limit districts' fiscal flexibility and ability to predict future allocations accurately.

The combination of these institutional features - the two-step allocation process, multiple formula thresholds, and annual poverty estimate updates - provides both the institutional context and empirical strategy for analyzing how Title I funding affects local education finance decisions in the current policy environment.

2.4 Data

My analysis combines several data sources at the school district level. The primary dataset for this study comes from the Department of Education's Title I Part A Allocation data, which contains detailed information on Title I. I also use information on the district characteristics and finance survey and tax and bond election data from Abott et al. (2020). I describe the datasets and variables in more detail.

2.4.1 Title I Allocation Information

The Department of Education in the United States provides allocation of Title I and detailed information to state education agencies when the allocation for the next academic year is determined. I use fiscal year revenue files for Title I

allocation, a district-level dataset which include final fiscal allocation of Title I to local education agencies by grant type, their hold-harmless base for each grant, and total Title I amount. The files also include formula counts used to determine Title I allocation, containing total formula count, 5-17 population, proportion of formula children compared to 5-17 population, eligible children for each grant, and weighted counts for Targeted grants and EFIG. From this dataset, I construct binary variables for each grant’s eligibility and the highest weights applied to districts, following the official calculation process of the Title I funding formula. The study spans the 2008-2009 through 2017-2018 academic years, comprising 136,547 observations with approximately 13,000 observations per year.

A key advantage of this dataset is that it provides the exact number of Title I formula counts at the district level, a crucial variable that has rarely been utilized in empirical research with the exception of Gordon and Reber (2023). This dataset uniquely enables a national-scale regression discontinuity design for several reasons. First, it provides the precise formula counts used in actual Title I allocation calculations, eliminating measurement error that would otherwise compromise identification. Second, the formula counts incorporate not only 5-17 children under poverty from SAIPE data, but also children who are neglected, delinquent, or in foster care—components that are typically unavailable at the district level from other sources. These additional components present significant data collection challenges: they come from different institutions with varying collection timelines, are often published only at aggregated state levels, and require de-duplication procedures to address potential double-counting across different administrative datasets. By providing these pre-calculated, official formula counts, this dataset overcomes these limitations and enables clean identification at formula thresholds. Additionally, the dataset’s annual updates allow for analysis of year-to-year changes in district eligibility and funding, providing variation that strengthens the identification strategy.

2.4.2 District Finances, Enrollment, and Demographic Information

I connect the Title I Part A Allocation data with multiple public datasets from the National Center of Education Statistics (NCES) Common Core of Data (CCD). The CCD provides two essential sources of information: the Local Education Agency Finance Survey (F-33) and the Local Education Agency Universe Survey, commonly referred to as the nonfiscal data.

The F-33 survey offers comprehensive financial information for school districts across the country. From this survey, I extract revenue data by source—federal, state, and local—with federal revenue further disaggregated to separate Title I from other federal programs. For expenditure categories, I focus specifically on four key measures: current elementary and secondary education expenditure (representing total current spending), capital outlays (covering construction, land acquisition, and equipment), current services expenditure (including student support services, instructional staff support, and administration), and current instructional expenditure. To ensure comparability across year, I adjust all financial variables to constant 2017 school year dollars using the `cpiget` Stata command by Shores and Candelaria (2020), which is specifically calibrated for education finance research.¹ This adjustment ensures that financial comparisons across different academic years reflect real changes rather than nominal inflation effects. I then normalize all financial variables on a per-pupil basis to facilitate meaningful comparison across districts of varying sizes.

The CCD’s annual nonfiscal survey complements the financial data with detailed demographic and enrollment information. I use this survey to obtain total

¹ The `cpiget` Stata command (Shores & Candelaria, 2020) provides education-specific inflation adjustments that align with academic year cycles rather than calendar years and accounts for the unique cost structures in educational institutions. Unlike standard CPI adjustments, this approach better reflects the purchasing power of education dollars across different time periods.

student enrollment (fall membership) and student demographic composition, including the percentage of students by race and ethnicity (White, Black, Hispanic, and other), as well as the proportion of English Language Learners (ELL) and special education students.

After merging with CCD data, my sample contains 134,437 observations, representing a reduction of 2,110 observations (1.5% of the original sample). This reduction primarily stems from excluding certain non-typical public school districts—districts that receive Title I funds but are organized under specialized or alternative administrative structures, rather than functioning as conventional K-12 school districts. To ensure data quality, I further exclude districts with missing values for total enrollment and key finance variables, resulting in a final analytic sample of 118,124 observations, with approximately 11,000 observations per year.

The merging process presented some challenges due to inconsistent district identifiers between the Title I allocation files and CCD datasets. While both sources use NCES district identification codes in principle, I discovered numerous instances where the same district had different codes across the two datasets. To address this issue, I employed probabilistic record linkage using Stata’s `reclink` command (Blasnik, 2010). This approach successfully resolved identifier mismatches for 166 district-year observations. For New York City’s public school district, which appears as five separate county-based observations in Title I allocation files² but as a single district in CCD, I manually aggregated these separate allocations into one consolidated NYC district observation to ensure proper merging with the financial data.

² New York, Bronx, Kings, Queens, and Richmond Counties

2.4.3 Tax and Bond Election Data

To explore the mechanism behind the observed decrease in capital expenditure, I incorporate data on local tax and bond elections from Abott et al. (2020). This dataset provides comprehensive information on closed bond and tax elections, allowing me to examine how Title I funding affects districts' local revenue-raising efforts. The tax election data covers 13 states³ from 1994 to 2017. The bond election data spans 15 states⁴ from 1993 – 2018. After merging with my dataset, the resulting observation is 30,867 for tax elections district-year observations in seven states (AR, LA, MI, MO, PA, TX, and WI) and 28,679 district-year observations for bond elections in six states (LA, MI, MO, PA, TX, and WI).

The election dataset provides information on whether districts held referenda in a given year (ever-had election) and whether these referenda passed (pass referenda), which are the two key outcome variables in my analysis of local revenue-raising efforts. While the original dataset contains additional details about specific election characteristics, such as dollar amount of bond issuance, and basis of proposed rate, I focus specifically on the binary outcomes of holding and passing elections as these directly address my research question regarding districts' responses to Title I funding.

A key advantage of this dataset is that it captures actual democratic decision-making processes regarding local education funding, rather than just expenditure outcomes. This allows me to examine whether Title I eligibility affects not only spending levels but also districts' propensity to seek additional local funding through democratic processes. The dataset's coverage period (1994-2017) fully encompasses my main analysis' timeframe, enabling me to observe concurrent changes in Title I

³ Arkansas, California, Louisiana, Michigan, Missouri, Nebraska, Nevada, Ohio, Oklahoma, Pennsylvania, Texas, Virginia, Wisconsin

⁴ Arkansas, California, Iowa, Louisiana, Maryland, Michigan, Missouri, Nebraska, Nevada, Ohio, Oklahoma, Pennsylvania, Texas, Virginia, Wisconsin

funding and local revenue-raising efforts.

[Table 2.2 about here]

One limitation is that the election data covers only a subset of states, potentially limiting generalizability. However, these states⁵ collectively enroll approximately 23% of U.S. public school students and represent diverse governance structures, funding formulas, and socioeconomic contexts. Additionally, these states have varying requirements for holding tax and bond elections, providing useful institutional variation for my analysis. The districts in the election sample show broadly similar characteristics to the full sample, though with some demographic differences (See Table 2.2)- notably in the proportion of Hispanic students (16% vs 14% in the full sample) and other racial groups (3.6% vs 7.4%). However, key fiscal variables like per-pupil revenues and expenditures remain comparable across samples, with differences well within standard deviations, suggesting the election samples maintain reasonable representativeness for analyzing district fiscal responses.

2.5 Empirical Strategy

To estimate the causal impacts of Title I on state and local revenue, I employ a stacked regression discontinuity design (RD) that leverages multiple discontinuities in the Title I funding formula. The key empirical challenge is isolating causal effects when funding is closely tied to district characteristics and child poverty especially. In addition, the complex eligibility structure of Title I creates several specific methodological challenges that require a specialized RD approach.

2.5.1 Endogeneity challenges in Title I analysis

The key empirical challenge in studying Title I's impact is that funding allocation is directly determined by district characteristics that may independently

⁵ Arkansas, Louisiana, Michigan, Missouri, Pennsylvania, Texas, Wisconsin

financial decisions. Title I funding depends primarily on two factors: (1) Title I formula count, the majority of them are from Small Area Income and Poverty Estimates (SAIPE), and (2) state per-pupil expenditure (SPPE). Both factors are likely correlated with unobserved district characteristics that influence financial outcomes.

A regression discontinuity design offers a solution by exploiting sharp eligibility thresholds in the Title I formula. By comparing districts just above and below these thresholds, I can identify causal effects under the assumption that districts near the cutoffs are similar in all respects except for their Title I eligibility status. However, Title I’s complex formula structure creates additional methodological challenges that require a specialized RD approach.

2.5.2 Identification Challenges in Multi-Threshold Title I Formula

Title I’s complex funding structure creates two significant estimation challenges that require a specialized approach beyond standard RD designs:

First, with multiple eligibility cutoffs and criteria, a traditional RD approach would produce several separate estimands—one for each threshold—with limited statistical power and no straightforward way to interpret these results collectively. Similar to challenges in difference-in-differences (DID) settings where individual group-by-time effects are less informative than their aggregation, Title I analysis benefits from consolidating information across thresholds. The stacked RD approach addresses this by combining evidence from all cutoffs into a single, interpretable coefficient while preserving the identification properties of the RD design.

Second, the overlapping nature of these thresholds creates potential contamination in the treatment-control distinction. Districts face different cutoffs for Basic Grants (2% poverty), Targeted Grants (5% poverty), and Concentration Grants (15% poverty), and a district exceeding the Basic Grant threshold may simultaneously exceed other thresholds. This creates a situation where “control” districts for higher-level thresholds may already receive treatment from lower-level grants,

violating the core assumption of clean separation between treatment and control groups.

2.5.3 Stacked Regression Discontinuity Approach

To address these challenges, I implement a stacked RD design that examines each cutoff separately while allowing for the estimation of an overall treatment effect. This approach allows for the identification of causal effects in the presence of multiple thresholds while accounting for threshold-specific heterogeneity in treatment effects.

The framework proceeds in three steps. First, I create separate datasets centered around each eligibility threshold, carefully selecting bandwidths that prevent contamination between different treatment conditions. Second, I combine these threshold-specific datasets vertically, maintaining their distinct identities through stack indicators. Finally, I estimate treatment effects while controlling for stack-specific characteristics that might influence district responses.

The stacked RD framework offers two key advantages for analyzing Title I's multi-threshold context. First, the framework maintains clean control groups within each threshold comparison. Rather than pooling all observations into a single RD—which would implicitly assume constant treatment effects—the stacked approach preserves the distinct nature of each threshold's quasi-experiment. This design is conceptually similar to the stacked difference-in-differences strategy employed by Cengiz et al. (2019), who construct separate control groups for each policy change to avoid contamination across treatment periods.

Second, by combining evidence across multiple thresholds, the stacked approach substantially increases statistical power while still allowing for the estimation of a meaningful overall treatment effect. This combined effect represents a weighted average of the threshold-specific effects, with weights determined by the precision of each threshold's estimate. In doing so, the approach aligns with recent methodological advances that emphasize careful handling of treatment effect heterogeneity

across multiple cutoffs (Cattaneo et al., 2016).

2.5.4 Bandwidth Selection Strategy

The complex structure of Title I eligibility required careful consideration when selecting appropriate bandwidths for the analysis. For example, Basic Grants and Targeted/EFIG Grants have threshold values at 2% and 5% poverty respectively, creating a situation where perfectly non-overlapping bandwidths would be impossible without severely restricting the data. After examining the distribution of districts across Title I eligibility thresholds, I determined that a uniform 3 percentage point bandwidth offered the optimal compromise.

This bandwidth choice provides sufficient statistical power across all analyses while maintaining relatively symmetric treatment of different cutoffs. More importantly, this data-informed approach effectively balances the competing objectives of maximizing available observations for statistical power and maintaining clean treatment-control comparisons, while minimizing data loss.

When this uniform approach created potential contamination between thresholds, I prioritized the Basic Grant cutoff in constructing comparison groups, as it represents the primary eligibility determination affecting the greatest number of districts. For instance, with the Basic Grant (2%) and Targeted/EFIG Grant (5%) thresholds in relatively close proximity, I limited the left-hand side bandwidth around Basic Grants to 2 percentage points to align with the formula threshold, meaning that all districts with less than 2% eligibility were included as controls. Similarly, I excluded observations around the 15.58% weighted count threshold to avoid contamination with the Concentration Grant (15%) analysis. This decision follows the previous literature in multi-threshold RD designs (Cattaneo et al., 2016; Eggers et al., 2018), where researchers maintain identification integrity by strategically defining comparison groups. For count criteria, I use a 100-student bandwidth, with the Basic Grant left-hand side restricted to 10 students.

[Table 2.3 about here]

The summary statistics for stacked samples are shown in Table 2.3. While schools in the count-based sample (column 1) tend to be smaller in terms of enrollment and number of Title I students, most fiscal and demographic characteristics remain comparable across samples. For instance, per-eligible Title I funding ranges from \$1,260 to \$1,320, and the proportion of Title I students varies between 17.3% and 20.6% across all samples. Core revenue sources and expenditure patterns also show similar levels, with differences well within their standard deviations. However, as shown in Figure 1, the distribution of Title I-eligible students is highly skewed - with a median of 200 students and 75th percentile of 550 students, only 1.8% of districts qualify for Concentration grants based on the count criteria, while 43% qualify through the proportion criteria. Given this skewed distribution in the count-based sample and its limited coverage of larger districts, as well as potential manipulation concerns revealed by McCrary density tests⁶, I proceed with the proportion-based sample as my main analytic sample.

[Figure 2.1 about here]

2.5.5 Estimation

The stacked RD approach builds upon the standard RD framework while addressing the challenges of multiple threshold. In a standard RD with a single cutoff, the treatment effect would be estimated as:

$$\delta = \lim_{\epsilon \rightarrow 0} E[Y_d | R_d = c + \epsilon] - \lim_{\epsilon \rightarrow 0} E[Y_d | R_d = c - \epsilon] \quad (2.1)$$

where Y_d represents the outcome for the district d , and R_d is the running variable — in this case, the number of Title I-eligible children or the proportion of such children

⁶ McCrary test results are discussed in Section 4.6

in the district’s 5-to-17-year-old population and c is the cutoff. When extending to multiple cutoffs, the stacked RD framework enables estimation across all thresholds while maintaining appropriate controls for threshold-specific characteristics.

Conceptually, the stacked RD estimate can be viewed as a pooled treatment effect across the various cutoffs $j = 1, \dots, J$. In a simplified illustration, If the policy effect is assumed homogeneous across all J thresholds, one can conceptually view the overall impact as an aggregation of these local discontinuities:

$$\delta_{\text{stacked}} = \sum_{j=1}^J w_j \cdot \lim_{\epsilon \rightarrow 0} \left(E[Y_d | R_{dj} = c_j + \epsilon] - \lim_{\epsilon \rightarrow 0} E[Y_d | R_{dj} = c_j - \epsilon] \right) \quad (2.2)$$

where w_j are weights indicating the relative contribution of each cutoff. A naive specification might set $w_j = N_j/N$, N_j denoting the sample size near cutoff j and N the total sample size across all stacks. However, this expression is primarily a conceptual simplification. In practice, I estimate δ_{stacked} by stacking the data from each cutoff into one regression—without explicitly imposing these weights—thereby letting ordinary least squares (OLS) internally determine how each cutoff influences the final coefficient.

I implement this framework through the following ordinary least square model. Specifically, I stack the data around each cutoff c_j , centering the running variable so that R_{dj} marks the threshold for stack j . I estimate a model of the form:

$$Y_{dtj} = \beta_0 + \beta_1 \text{Treat}_{dtj} + \beta_2 R_{dtj} + \beta_3 (\text{Treat}_{dtj} \times R_{dtj}) + X_{dtj} + \gamma_j (\pi_t + \sigma_s) + e_{dj} \quad (2.3)$$

where Y_{dtj} is the outcome for district d in year t , which belongs to cutoff j , and Treat_{dtj} indicates whether district d is above the cutoff in stack j in year t . R_{dtj} is the centered running variable for the district d in stack j at year t and X_{dtj} representing district covariates, which includes district demographic composition, percentage of special education students and English language learners. $\gamma_j, \pi_t, \sigma_s$

represent stack, year, and state fixed-effects, respectively. The key parameter of interest is β_1 , which captures the treatment effect across the cutoffs.

To account for heterogeneous treatment effects across cutoffs, I include stack-by-year and stack-by-state fixed effects. These capture variations in temporal shocks and regional characteristics specific to each threshold (Cengiz et al., 2019). While one could potentially add three-way (stack \times state \times year) fixed effects, doing so would leave minimal within-cell variation for identifying the RD effect, given the data constraints. Hence, I opt for the two-way interactions, which sufficiently address the primary sources of unobserved heterogeneity without unduly sacrificing statistical power.

2.5.6 Validating the Stacked RD Design

The credibility of the stacked RD design rests on two key identifying assumptions: the absence of manipulation at the eligibility thresholds and the comparability of districts just above and below these thresholds. I examine these assumptions through density tests and covariate balance tests, respectively.

[Table 2.4 about here]

Table 2.4 presents the McCrary density test throughout the thresholds in Title I formula for proportion-based sample (Cattaneo et al., 2020; McCrary, 2008). I apply the McCrary test to the raw running variable to verify that districts do not manipulate their Title I eligibility.⁷ Except for Threshold 4, there is no clear evidence of discontinuities, suggesting no systematic manipulation of Title I eligibility thresholds. At Threshold 4, the p-value of 0.046 indicates marginal significance, suggesting a possible discontinuity in density of districts. However, it is neither so large

⁷ While the main RD regressions include fixed effects (e.g., state-by-stack and year-by-stack) to control for unobserved heterogeneity, I conservatively apply the McCrary to the original, unadjusted distribution of the running variable. In practice, this means that cross-state differences in eligibility rates at the cutoff could exaggerate rejection rates.

nor so robust as to confirm a pronounced or systematic manipulation. Furthermore, because districts often lack precise year-to-year control over proportion of the Title I kids, any observed discontinuities are unlikely to reflect perfect sorting right at the cutoff. Lee and Lemieux (2010) mention that in the absence of tight, finely grained control, the near-cutoff variation resembles random assignment, upholding the RD identifying assumption.

[Figure 2.2 about here]

Figure 2.2 complements these tabular findings by visually depicting the distribution of districts around each proportion-based cutoff. Although the Basic threshold (2%) plot shows an overall upward slope, the test results in Table 2.4 confirm no statistically significant jump at the cutoff. Meanwhile, near Threshold 4 (38.24%), there appears to be a slight kink consistent with the marginally significant p-value (0.046). However, as discussed above, districts' lack of precise year-to-year control suggests that even this small irregularity does not reflect a systematic effort to manipulate Title I eligibility. Overall, Figure 2.2 reinforces the conclusion that no strong evidence of discontinuity emerges at any cutoff, thereby supporting the RD identifying assumptions.⁸

For completeness, Appendix Table A.1 and Figure A.1 replicate the McCrary test using the count-based Title I thresholds (e.g., at least 10 formula children, 6,500 for Concentration, etc.). Because over 75 percent of districts have fewer than 550 formula children, the higher count cutoffs yield extremely sparse observations, making the test either unstable or undefined. While the Basic Grant cutoff (10) shows a significant discontinuity at first glance, its effective sample is very different from the proportion-based thresholds, and the other count-based margins involve

⁸ Appendix Figure A.2 provides a visual check for potential manipulation of the running variable across all states, showing histograms of the re-centered proportion of eligible children for each state. The absence of visible clustering around the thresholds (marked by vertical red lines) further supports the validity of the RD design across geographic contexts.

too few districts to be analytically meaningful. Accordingly, I do not rely on the count-based sample for the main analysis, instead focusing on the proportion-based sample where the number of districts near each cutoff is sufficiently large and better suited for a robust RD design.

[Table 2.5 about here]

Next, I turn to a covariate balance test to verify the quasi-random assignment assumption, which requires that districts just above and below each cutoff be similar in observable characteristics. As shown in Table 2.5, these characteristics are statistically indistinguishable on either side of the threshold for all but one variable: the share of Hispanic students. While the Hispanic share is about 0.4 percentage points lower on the treated side, the difference is negligible relative to the 14% average proportion of Hispanic students in the sample.

This balance test is particularly important because I employ a stacked RD design without additional covariates, which yields more conservative estimates but hinges heavily on the assumption of local randomization near the cutoff. The absence of meaningful differences in other district characteristics thus supports the credibility of the RD framework, suggesting no systematic sorting that would violate the quasi-random assignment at these thresholds.

2.6 Results

2.6.1 First Stage Results: Change in Title I Allocation

The identification strategy leverages discontinuities in the Title I funding formula to estimate the causal effects of federal aid on local fiscal behavior. Table 2.6 presents first-stage estimates based on equation (3).⁹ At the formula cutoffs,

⁹ In column (2), I estimate the model using Poisson pseudo-maximum likelihood (PPML) following Silva and Teneyro (2006) and convert the coefficient to a percentage effects as explained in Chen and Roth (2024). This approach provides a proper percentage interpretation that is invariant to the scaling of the outcome variable. The estimation was implemented using Correia et al. (2019)'s `ppmlhdfc` Stata package.

Title I allocation increases by \$64.68 per eligible child (SE: 5.840), representing a 4.7 percent rise in funding (SE: 0.012). For districts with average formula children counts (864.6 children), this translates to approximately \$56,000 in additional Title I funding.¹⁰

[Table 2.6 about here]

The precision of these estimates, combined with the McCrary (Table 2.4) and covariate balance tests (Table 2.5), provide credible evidence that these cutoffs are both exogenous and sufficiently powered to distinguish behavioral differences between eligible and non-eligible districts that examining state and local responses to Title I eligibility can be fruitful. The magnitude and statistical significance of these first-stage effects support interpreting the reduced-form estimates as the causal effects of Title I on district financial behavior.

The first stage of my analysis focuses on Title I allocation per eligible child, consistent with the program's statutory design. Because Title I is awarded based on the number (or proportion) of low-income students, denominating dollars this way provides a direct indication of how much additional federal aid flows to eligible students in districts meeting the funding formula criteria. However, in examining the second stage, I use per-pupil revenue and spending measures denominated by total district enrollment. This approach is more appropriate for measuring overall district financial impacts because denominating total revenues by eligible children alone would artificially inflate per-student figures in districts with lower proportions of

¹⁰ Appendix Table A.2 separates the analysis by Basic and Non-Basic grants, showing that the funding discontinuity is substantially larger for Basic grants (\$896 per eligible child or 233.5% increase) compared to Non-Basic grants (\$38 per eligible child or 2.8% increase). This is consistent with the formula structure where Basic grants provide the foundation of Title I funding, while Non-Basic grants serve as targeted supplements. Despite these differences in magnitude, the overall pattern of fiscal responses remains similar across grant types, supporting the pooled analysis approach in the main text.

eligible students, just as, conversely, denominating Title I funds by total enrollment would understate the actual aid directed to eligible students.

2.6.2 Effects on Districts' Finance

Table 2.7 provides evidence that the relationship between Title I funding and local fiscal behavior has evolved since the early implementation period. Contrary to previous studies that found substantial crowd-out effects, I find no evidence that marginal increases in Title I funding lead to reductions in state or local education revenue. The coefficient on per-pupil state revenue is \$75.7 (SE=61.7), suggesting that state funding tends to complement rather than offset federal aid, though this increase is not statistically significant. Local revenue shows an even larger complementary response, with an estimated increase of \$182.29 (SE=125.9) at the cutoff. While the point estimates suggest complementary effects, the wide confidence intervals prevent strong conclusions about the extent of fiscal substitution. The lower bound of the 95% confidence interval for local revenue (approximately -\$70) would be consistent with substantial crowd-out, potentially offsetting much of the Title I increase of \$64.68 per eligible child. Conversely, the upper bound would suggest significant complementary effects. This highlights the statistical uncertainty in precisely quantifying the fiscal response, though the point estimates consistently point toward non-negative fiscal responses.^{11 12}

¹¹ Table A.3 reports Poisson pseudo-maximum likelihood (PPML) estimates, translating the effects in Table 2.7 into approximate percentage changes. Consistent with the OLS results, capital outlays show a significant 6.4% decrease at the Title I margin, while all other expenditure categories exhibit small (1–2%) and statistically insignificant changes. This reinforces the finding that Title I funding primarily displaces capital investments rather than day-to-day spending.

¹² The coefficient on total federal revenue appears modest and slightly negative (-4.035, SE=16.26), despite the increase in Title I funding shown in the first-stage results. This pattern appears to be driven by small offsetting changes in other federal programs, particularly Child Nutrition Act funding, which shows a similar magnitude decrease (-3.37, p<0.1) at Title I formula cutoffs. These offsetting effects are only marginally significant and do not affect the main conclusions about state and local responses to Title I funding. (See Table A.4.)

[Table 2.7 about here]

These findings stand in sharp contrast to much of the pre-2000s literature, which frequently reported sizable or even near-complete crowd-out (Cascio et al., 2013; Gordon, 2004). In that earlier period, local governments often reduced their own contributions to K-12 education after receiving new federal aid, negating the intended additive effect of Title I. The present results suggest that, in the post-adequacy era, districts retain most of the incoming federal dollars in their operating budgets. These patterns could reflect various institutional changes (maintenance-of-effort requirements, new accountability systems, annual SAIPE poverty updates, etc.), which may limit local offset. Section 5.4 examines whether state-level policy variation—particularly funding progressivity—helps explain these shifts.

While operational expenditures remain effectively insulated from crowd-out, capital expenditures decline at Title I margins, pointing to an indirect form of budget adjustment. Districts at Title I margins show no significant changes in core spending areas—elementary/secondary expenditures (\$155.01, SE=103.28), student services (\$56.89, SE=42.05), and instruction (\$104.88, SE=63.83). However, capital expenditures show a significant decline of \$83.49 per pupil (SE=48.23) at Title I margins. Unlike operational spending, which faces stricter rules and higher political visibility, capital projects appear more vulnerable to being scaled back when federal funds are available.

2.6.3 Effects on Local Tax and Bond Elections

To investigate this selective reduction in capital spending, I turn to tax and bond election outcomes. I use closed tax and bond election data from Abbott et al.

(2020) and estimated equation (3).¹³ Table 2.8 presents two sets of outcomes—whether districts held elections and whether they pass referenda in a given year (or average referenda pass rate if the district has more than one election in a given year). The results show interesting differences between tax and bond elections. Title I eligibility has no significant effect on either holding (-0.002, SE=0.013) or passing (0.001, SE=0.007) tax referenda, which primarily fund operational expenses. In contrast, districts are 2.8 percentage points less likely to hold bond elections at Title I margins (SE=0.012), though when they do hold elections, they show slightly higher passage rates (1.6 percentage points, SE=0.007). Thus, even if day-to-day budgets are supplemented, districts offset some portion of Title I aid by postponing or reducing large-scale capital projects, effectively lowering the likelihood of incurring new debt obligations.

[Table 2.8 about here]

This pattern becomes even more pronounced when examining districts by their Title I funding levels.¹⁴ Table 2.9 shows that districts in the highest quartile of Title I funding (within their state-year) are 4.7 percentage points less likely to hold bond elections (SE=0.023), while those in lower quartiles show no significant changes. These results suggest that while districts maintain their operational

¹³ Specifically, I set the outcome to be 1 if a tax (or bond) election was attempted and 0 if not; for states without data, it is set to missing. Separately, I also set the outcome to be 1 if a tax (or bond) election was passed and 0 if lost; for states without data, it is set to missing, whereas for districts without elections it is set to zero, and I include indicator variables set to 1 if those districts did not have elections.

¹⁴ I classify districts into quartiles based on Title I amount per eligible child within each state-year using the full sample of all U.S. school districts. These quartile assignments are created using the original, unstacked dataset to preserve the natural distribution of Title I funding across districts. When these quartile indicators are merged with the stacked RD sample and subsequently with Abott et al. (2020)'s tax and bond election data, the resulting number of observations in each quartile differs from perfect quarters, as shown in Table 2.9. This uneven distribution reflects that the quartile assignment was based on the complete universe of districts, not just those near the formula cutoffs in the stacked analysis sample.

spending levels under Title I, they may offset some portion of federal aid by reducing capital investments, particularly in areas receiving larger Title I allocations.

[Table 2.9 about here]

2.6.4 Implication on State Funding Policies

Having established that districts reduce capital investments rather than operating funds in response to Title I, a natural question is whether this behavior varies with state institutional features. The institutional changes of the adequacy era, particularly state-level adequacy reforms and funding constraints, might create different environments for districts' responses to federal aid. For example, states with more progressive funding formulas or stricter property tax regulations could either amplify or constrain districts' ability to adjust their capital spending.

[Table 2.10 about here]

To investigate this possibility, I examine whether state-level funding progressivity predicts the magnitude of district responses to Title I (see Table 2.10).¹⁵ I also

¹⁵ I estimate state-specific responses to Title I by running equation (3) for separately for all states, controlling student demographics and year-by-stack fixed effects ($\widehat{\beta}_{1s}$). Then I construct measures of state-level funding progressivity for each revenue source (state, local) and expenditure categories. I define progressivity as the enrollment-weighted mean differences in per-pupil funding between Title I eligible and non-eligible students, following previous literature (Chingos & Blagg, 2017; Knight & Mendoza, 2019; Lee et al., 2022). For each state s , I estimate the following equation:

$$Y_{dt} = \alpha_{0s} + \alpha_{1s}TitleI_{dt} + \pi_t + \epsilon_{dt} \text{ (for district } d, \text{ in state } s)$$

where Y_{dt} is per-pupil funding in district d in year t , $TitleI_{dt}$ indicates Title I eligibility status and the model includes year fixed effects (π_t). The regression is weighted by the number of Title I eligible and non-eligible kids in each district. The coefficient of interest, α_{1s} , captures the weighted mean funding difference between Title I and non-Title I students in state s . Equivalently,

$$\widehat{a}_{1s} = \frac{\sum_{d \in s,t} w_{dt} (Y_{dt}^{TitleI=1} - Y_{dt}^{TitleI=0})}{\sum_{d \in s,t} w_{dt}}$$

where w_{dt} is the enrollment in district d in year t , used as weight. All other notations remain as previously defined. The funding progressivity measure $\widehat{\alpha}_{1s}$ is calculated for state revenue, local revenue and elementary/secondary expenditures. I then merge these state-specific estimates with the state-level progressivity measures ($\widehat{\alpha}_{1s}$) and conducted a meta-regression as follows:

incorporate two additional indicators from EdBuild’s “FundEd” data—property tax floors and ceilings—since statutory minimum or maximum tax rates could constrain districts’ ability to raise or lower local revenue in response to federal grants.¹⁶

Table 2.10 presents meta-regression results that explore how state funding characteristics relate to Title I effects on district finances. The progressivity measures capture the extent to which state funding systems direct additional resources to disadvantaged students, calculated as the enrollment-weighted mean differences in per-pupil funding between Title I eligible and non-eligible students. This approach follows previous literature on measuring funding progressivity (Chingos & Blagg, 2017; Knight & Mendoza, 2019; Lee et al., 2022).

The results reveal several important patterns. First, in states with higher elementary/secondary expenditure progressivity, the negative effects of Title I on capital outlays are significantly stronger (-0.157, $p < 0.05$). To interpret this coefficient, we need to understand that the progressivity measure captures how much additional funding disadvantaged students receive compared to non-disadvantaged students within a state. The coefficient of -0.157 indicates that for each additional \$1 in elementary/secondary expenditure progressivity, the effect of Title I on capital outlays decreases by \$0.157. In simpler terms, districts in states that direct more funding to disadvantaged students respond to Title I funding by making larger reductions in their capital expenditures. This suggests that when state funding systems already concentrate resources on disadvantaged students, districts appear to have greater flexibility (or perhaps incentive) to reduce capital investments in response to Title I increases.

$$\widehat{\beta}_{1s} = \delta_0 + \delta_1 \widehat{\alpha}_{1s} + \mu_s$$

where $(\widehat{\beta}_{1s})$ denotes the Title I effect in state s , and $(\widehat{\alpha}_{1s})$ is that state’s progressivity measure. The observations are weighted by the inverse of their estimated variances.

¹⁶ See EBuild, FundEd - Property Tax Bound for details. (<http://funded.edbuild.org/national#property-tax-bounds>)

Interestingly, state revenue progressivity does not significantly predict Title I's effects on any district finance outcomes. The coefficient for state revenue progressivity on capital outlays is negative (-0.035) but not statistically significant. This means that for each additional \$1 in state revenue progressivity, there is a non-significant \$0.035 reduction in the Title I effect on capital outlays per pupil. This suggests that the progressivity of expenditure mechanisms rather than the progressivity of revenue distribution itself is what moderates districts' capital spending responses. In practical terms, how states actually spend their money matters more than where they get it from when it comes to influencing district responses to Title I.

Moreover, neither property tax floor nor property tax ceiling provisions exhibit a meaningful association with how districts respond to Title I. For property tax floors, the coefficient is -\$21.27 on state revenue and \$80.25 on capital outlays, while for property tax ceilings, the coefficients are \$1.13 on state revenue and -\$0.30 on capital outlays. All these coefficients are statistically insignificant, with standard errors exceeding the point estimates in all specifications. Given that these are binary indicators (1 if the state has the provision, 0 if not), the interpretation would be that having a property tax floor is associated with a non-significant \$80.25 increase in the Title I effect on capital outlays per pupil. The large magnitude but statistical insignificance of these estimates indicates substantial variation in how districts respond to Title I within states that have similar tax policies. This finding suggests that while states impose various property tax rules, these constraints do not systematically influence the capital-spending adjustment documented above.

These results have important implications for understanding how state policy environments interact with federal grants. While the adequacy-era reforms have, at a national level, curtailed the crowd-out documented in earlier decades (Cascio et al., 2013; Gordon, 2004), variation in state progressivity does moderate districts' responses to Title I. Indeed, the relationship between state elementary/secondary

expenditure progressivity and Title I-induced capital spending reductions suggests that state funding structures play a role in shaping how districts reallocate resources in response to federal aid.

However, the lack of systematic relationships between most state progressivity measures and Title I outcomes suggests that the capital-spending adjustment mechanism identified in this paper operates across diverse state funding environments and is not easily predictable by conventional school funding policy characteristics. This consistency reinforces the robustness of the main findings and indicates that districts retain sufficient autonomy to strategically scale back long-term projects or bond issuances when federal aid increases, regardless of specific state policy nuances.

The overall shift toward more supportive funding structures appears to have established a baseline environment in which Title I mainly supplements operating resources across states, regardless of individual policy details. Nonetheless, capital investments remain a channel for local adjustments under all these varied state contexts, enabling districts to subtly reallocate or delay infrastructure spending despite broad gains in equity-oriented school finance policy. This finding underscores the importance of considering multiple dimensions of fiscal responses when evaluating the effectiveness of intergovernmental grants like Title I.

2.7 Discussion and Conclusion

This study documents a fundamental shift in how local governments respond to Title I funding in the post-adequacy era. While earlier research conducted in different institutional contexts found substantial crowd-out effects (Cascio et al., 2013; Gordon, 2004), my findings suggest that Title I no longer substitutes for operational funding but instead influences capital investment decisions.

This shift reflects two key institutional mechanisms. First, the increasing progressivity of state funding formulas and stricter maintenance-of-effort provisions

have constrained districts' ability to offset federal aid through reductions in operational budgets (Candelaria & Shores, 2019; Lafortune et al., 2018). Second, the transition to annual SAIPE poverty estimates has introduced more frequent and less predictable funding adjustments, further complicating strategic fiscal planning.

These constraints have not eliminated fiscal substitution but rather redirected it toward capital spending adjustments. Districts at Title I margins reduce capital outlays by \$83 per pupil and are 2.8 percentage points less likely to hold bond elections, with the strongest effects concentrated in high-Title I districts. These patterns align with tax salience theory (Cabral & Hoxby, 2012), which suggests that policymakers should expect local adjustments to emerge in less visible tax channels when regulatory constraints limit operational budget flexibility.

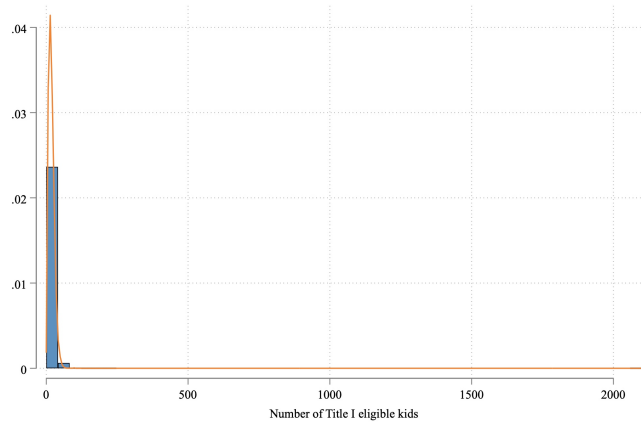
These findings contribute to the broader literature on intergovernmental grants and local fiscal behavior. Prior research has debated whether federal grants increases education spending (Card & Payne, 2002; Hoxby, 2001; Inman, 2008), or merely substitute for local funding (Cascio et al., 2013; Gordon, 2004; Nguyen-Hoang & Yinger, 2020). Earlier studies found that categorical education grants often result in local tax reductions rather than net increases in school funding (Fisher & Papke, 2000; Steinberg et al., 2016), and that school finance equalization efforts sometimes lead to shifts in local revenue allocation rather than increased investment in education (Baicker & Gordon, 2004; Chakrabarti & Setren, 2011). However, unlike prior findings suggesting that localities primarily substitute federal aid through tax adjustments, my results show a shift toward capital spending reductions, marking a distinct form of fiscal substitution in the post-adequacy era.

Although further work is needed to examine the long-term consequences of these adjustments, my findings suggest that federal education grants must be designed with safeguards to prevent unintended capital spending reductions. Policymakers should consider expanding infrastructure-specific funding streams or incorporating targeted capital investment provisions into Title I to ensure that federal

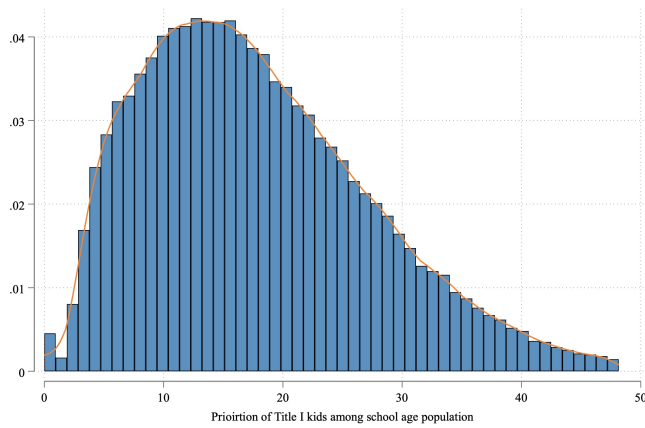
aid enhances long-term educational resources rather than indirectly displacing infrastructure investment.

Figure 2.1: Distribution of Running Variables

(a) Title I Eligible Kids

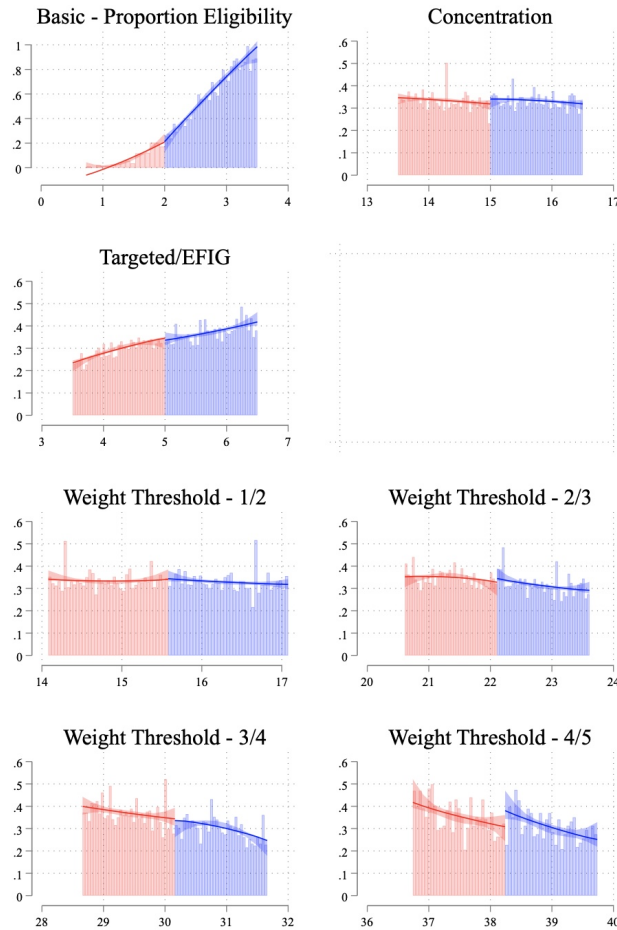


(b) Proportion of Title I Eligible Kids



Note: Panel (a) shows the histogram of the number of Title I eligible children in each district, and Panel (b) shows the histogram of the proportion of Title I eligible children among the 5–17 population. The y-axis represents the density of each distribution, and the orange line indicates the kernel density estimate. The sample includes all districts receiving Title I from the 2008–09 through 2017–18 academic years.

Figure 2.2: Density of observations (proportion criteria)



Note: Each panel displays the distribution of the running variable around the indicated Title I grant/weight cutoff (vertical boundary). Red bars represent observations below the cutoff; blue bars are those above. The red/blue curves are local polynomial density estimates from `rddensity` using a polynomial order of 2 and a bias-correction order of 3, with a bandwidth of ± 3 (except for the Basic Grant, where the left bandwidth is 2). A visible jump at the cutoff indicates potential sorting or manipulation of observations.

Table 2.1: Title I Grant Eligibility and Weighting Formula

Eligibility Requirements for Title I Grants		
Grants	Requirements	
Basic	The district must have more than 10 formula children “AND” the number of those children must exceed 2% of the total 5-to-17 population of the districts.	
Concentrated	The district must have at least 6,500 formula children “OR” the number of those children must be or exceed 15% of the total 5-to-17 population of the districts.	
Targeted and EFIG	The district must have more than 10 formula children “AND” the number of those children must exceed 5% of the total 5-to-17 population of the districts.	
Weighting Formula for Targeted Grants and EFIG		
Weight	Formula Count	Proportion
1	Up to 691	Up to 15.58%
1.5	692 – 2,262	15.58% - 22.11%
2	2,263 – 7,851	22.11% to 30.16%
2.5	7,852 – 35,514	30.16% to 38.24%
3	Over 35,515	Over 38.24%

Notes: Title I eligibility requires thresholds based on the number and proportion of formula children in a school district. Weighting increases for higher formula child counts and proportions, enhancing the funding calculation for Targeted Grants and EFIG.

Table 2.2: District Fiscal and Nonfiscal Characteristics (Summary Statistics)

	Full Sample (1)	Tax Election (2)	Bond Election (3)
Per Eligible Child Title I	1316.3 (587.0)	1262.0 (363.2)	1256.0 (367.8)
Per Pupil Title I	284.8 (546.6)	296.1 (272.5)	289.1 (274.1)
Panel A: District Revenue and Expenditures			
Total Revenue	15623.2 (12595.9)	14225.5 (7341.8)	14429.8 (7561.9)
Federal Revenue	1267.1 (1630.6)	1210.0 (958.5)	1184.8 (967.2)
State Revenue	6940.2 (5607.1)	5972.6 (2500.6)	5955.6 (2560.1)
Local Revenue	7415.9 (9773.5)	7042.9 (7609.0)	7289.3 (7820.0)
Current Elementary/Secondary Expenditure	12982.2 (7884.2)	11456.9 (2987.8)	11562.0 (3050.5)
Current Instructional Expenditure	7751.8 (4596.4)	6860.5 (1721.5)	6946.1 (1746.2)
Current Service Expenditure	4683.3 (3693.5)	4052.6 (1416.1)	4076.5 (1450.3)
Capital Outlays	1293.5 (3078.8)	1328.3 (2915.4)	1342.4 (2984.8)
Fall Membership	3808.9 (13125.5)	3459.9 (9049.1)	3562.3 (9338.4)
Panel B: District Demographics			
Number of Title I Formula Children	796.2 (3841.4)	788.2 (3314.2)	806.8 (3429.7)
Prop. of Title I Formula Children (/5-17 population)	0.180 (0.101)	0.198 (0.100)	0.193 (0.0999)
Prop. of White Students	0.718 (0.272)	0.727 (0.269)	0.726 (0.270)
Prop. of Black Students	0.0719 (0.153)	0.0785 (0.155)	0.0722 (0.144)
Prop. of Hispanic Students	0.137 (0.203)	0.157 (0.226)	0.163 (0.232)
Prop. of Other Race Students	0.0737 (0.127)	0.0379 (0.0516)	0.0385 (0.0529)
Prop. of ELL Students	0.0476 (0.0913)	0.0399 (0.0693)	0.0403 (0.0693)
Prop. of Special Education Students	0.139 (0.0522)	0.128 (0.0440)	0.128 (0.0450)
Ever Had Tax/Bond Election in the Year	–	0.344 (0.475)	0.411 (0.492)
Average Pass Rate of Referenda	–	0.744 (0.379)	0.715 (0.365)
Observations	118,124	30,867	28,679

Notes: Descriptive statistics in this table are based on districts receiving Title I revenue from the 2008–09 through 2017–18 academic years. Districts that receive Title I but are not included in the Common Core of Data are excluded from the sample. All dollar amounts are adjusted to 2017 dollars. Columns (2) and (3) use local tax (1994–2017) and bond (1993–2018) election data from Abott et al. (2020). After merging with the main dataset, these datasets cover 30,867 district-year observations across seven states for tax elections and 28,679 district-year observations across six states for bond elections, which explains the difference in sample sizes. The reported means are unweighted. When weighted by district enrollment, the proportions of White, Black, and Hispanic students are 52%, 15.3%, and 24.1%, respectively, which aligns more closely with the national average. (Mean values; SD in parentheses)

Table 2.3: District Fiscal and Nonfiscal Characteristics (Stacked Sample)

	Count Stack (1)	Proportion Stack (2)	Tax Election (3)	Bond Election (4)
(ED) per-eligible child Title I	1320.3 (636.2)	1305.0 (533.0)	1266.2 (343.1)	1260.8 (348.2)
(ED) per-pupil Title I	274.0 (378.4)	302.6 (336.3)	309.0 (234.7)	304.1 (238.0)
Panel A: District Revenue and Expenditures				
Total Revenue	16552.3 (16368.0)	15342.9 (12708.8)	14135.5 (7840.0)	14376.6 (8126.4)
Federal Revenue	1297.5 (1712.4)	1332.2 (1514.7)	1264.9 (938.9)	1244.6 (956.3)
State Revenue	7104.8 (7063.1)	7038.6 (6069.0)	5997.7 (2485.3)	5984.1 (2558.2)
Local Revenue	8150.0 (12511.0)	6972.1 (9458.6)	6872.9 (8093.1)	7147.9 (8373.5)
Current Elementary/Secondary Expenditure	13589.7 (9786.1)	12799.9 (8346.4)	11373.1 (2969.5)	11497.6 (3046.6)
Current Instructional Expenditure	8048.8 (5490.3)	7624.9 (4912.9)	6794.2 (1683.5)	6892.8 (1713.6)
Current Service Expenditure	4966.4 (4773.9)	4612.8 (3633.7)	4022.2 (1424.2)	4051.7 (1467.2)
Capital Outlays	1333.4 (3596.2)	1266.2 (3065.4)	1313.4 (2889.9)	1328.2 (2969.4)
Panel B: District Demographics				
Fall Enrollment	2363.6 (7196.0)	3795.2 (13404.6)	3228.0 (9196.8)	3330.4 (9549.5)
Number of Title I Formula Children	447.1 (1780.2)	865.0 (3986.7)	789.0 (3386.1)	812.5 (3525.7)
Prop. of Title I Formula Children (/5–17 population)	0.173 (0.0995)	0.195 (0.0854)	0.210 (0.0791)	0.206 (0.0787)
Prop. of White Students	0.742 (0.259)	0.707 (0.275)	0.722 (0.267)	0.719 (0.270)
Prop. of Black Students	0.0611 (0.143)	0.0746 (0.153)	0.0796 (0.152)	0.0739 (0.143)
Prop. of Hispanic Students	0.123 (0.187)	0.145 (0.210)	0.162 (0.224)	0.170 (0.231)
Prop. of Other Race Students	0.0737 (0.134)	0.0740 (0.127)	0.0364 (0.0490)	0.0370 (0.0505)
Prop. of ELL Students	0.0423 (0.0864)	0.0494 (0.0924)	0.0393 (0.0645)	0.0397 (0.0639)
Prop. of Special Education Students	0.138 (0.0578)	0.140 (0.0519)	0.128 (0.0442)	0.129 (0.0454)
Ever Had Tax/Bond Election in the Year	–	–	0.336 (0.472)	0.403 (0.491)
Average Pass Rate of Referenda	–	–	0.201 (0.386)	0.262 (0.412)
Observations	61,968	72,104	19,813	18,175

Notes: Descriptive statistics in this table are based on districts receiving Title I revenue from the 2008–09 through 2017–18 academic years. Districts that receive Title I but are not included in the Common Core of Data are excluded from the sample. All dollar amounts are adjusted to 2017 dollars. Columns (1) and (2) present summary statistics for the stacked dataset, using the count and proportion criteria, respectively. The proportion-based stack is our main analytic sample. Columns (3) and (4) incorporate local tax and bond election data from Abott et al. (2020), merged with the proportion-based sample. The reported means are unweighted. (Mean coefficients; SD in parentheses)

Table 2.4: McCrary Density Test

	Weight	Transition	Cutoff	Obs Left	Obs Right	T-statistics	p-value	BW Left	BW Right
Basic	-	-	2	323	6768	-0.744	0.457	2	3
Concentration Targeted/EFIG	-	-	15	14642	14144	1.052	0.293	3	3
Threshold 1	1 → 1.75	-	5	6743	11829	-0.664	0.507	3	3
Threshold 2	1.75 → 2.5	15.58	22.11	14577	13805	0.878	0.380	3	3
Threshold 3	2.5 → 3.25	22.11	30.16	11380	9622	0.349	0.727	3	3
Threshold 4	3.25 → 4	30.16	38.24	6485	4571	0.284	0.776	3	3
				2551	1714	1.991	0.046	3	3

Notes: This table reports results from McCrary-type density tests (Cattaneo et al., 2020; McCrary, 2008) on the proportion-based sample. “Obs Left” and “Obs Right” denote the number of districts within the chosen bandwidth on each side of the cutoff. Failing to reject the null ($p > 0.05$) implies no significant manipulation at the threshold. “BW Left” and “BW Right” show the local polynomial bandwidths used on each side. All tests use a triangular kernel, polynomial order $p = 2$, bias order $q = 3$, and a chosen bandwidth of 3 (except BW Left = 2 for Basic grant).

Table 2.5: Covariate Balance Test

Covariate Balance Test							
(1)	(2)		(3)		(4)		(5)
	Black	Hispanic	White	Other Race	ELL	Special ED	Total Enrollment
Formula Cutoff	-0.001 (0.001)	-0.004** (0.002)	0.003 (0.003)	0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-345.390 (275.260)
Proportion of Title I Kids	0.005*** (0.001)	0.006*** (0.002)	-0.012*** (0.002)	0.002* (0.001)	0.002*** (0.001)	0.001*** (0.000)	84.019 (107.305)
Cutoff x Proportion of Title I Kids	-0.000 (0.001)	0.001 (0.001)	0.002 (0.002)	-0.002* (0.001)	0.000 (0.001)	-0.000 (0.000)	-34.055 (105.495)
Observations	72,104	72,104	72,104	72,104	72,104	72,104	72,104
Mean	0.0746	0.145	0.707	0.0740	0.0494	0.140	3794.1
R ²	0.445	0.535	0.488	0.445	0.338	0.327	0.132
							864.6
							0.0925

Notes: Estimates use the indicated student covariate as the dependent variable, controlling for stack-by-year and stack-by-state fixed effects. The sample is the proportion-based stack, which is our main analytic sample. I employ a stacked regression discontinuity (RD) design without other district covariates, producing more conservative estimates. Standard errors are clustered at the state level, with robust standard errors in parentheses. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.)

Table 2.6: First Stage: Effects of Title I Formula Cutoffs on Title I Revenues

	(1) FE Estimates	(2) PML Estimates
Formula Cutoff	64.681*** (5.840)	0.047*** (0.012)
Proportion of Title I Kids	11.940*** (2.381)	0.008*** (0.003)
Cutoff \times Proportion	2.921 (3.490)	0.004 (0.003)
Observations	72,102	72,102
Mean	1305.0	1305.0
R² (Pseudo R²)	0.565	0.497

Notes: This table reports estimates from a stacked regression discontinuity (RD) design using data from the 2008–09 through 2017–18 academic years. Column (1) shows results from an OLS model with fixed effects, and Column (2) uses a Poisson pseudo–maximum likelihood estimation using Stata `ppmlhdfe` command (Correia et al., 2019). All dollar amounts (per-eligible-child Title I) are adjusted to 2017 dollars. Both models include controls for district demographics (including racial composition, special education and English learner proportions), as well as stack-by-state and stack-by-year fixed effects. Standard errors are clustered at the state level, with robust standard errors in parentheses. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.)

Table 2.7: Effects on District Finance (Per Pupil)

	(1)	(2)	(3)	(4)
Panel A: Revenues				
	Total Revenue	Federal Revenue	State Revenue	Local Revenue
Formula Cutoff	253.973 (155.412)	-4.035 (16.260)	75.715 (61.721)	182.293 (125.972)
N	72,102	72,102	72,102	72,102
Mean	15,342.8	1,332.1	7,038.4	6,972.3
R²	0.123	0.442	0.219	0.178
Panel B: Expenditures				
	Elem/Sec Exp	Capital Outlays	Services	Instruction
Formula Cutoff	155.011 (103.278)	-83.486* (48.233)	56.894 (42.051)	104.880 (63.829)
N	72,102	72,102	72,102	72,102
Mean	12,799.8	1,266.1	4,612.7	7,624.8
R²	0.192	0.0396	0.158	0.220

Notes: This table reports estimates from a stacked regression discontinuity (RD) design using data from the 2008–09 through 2017–18 academic years. All models use OLS with fixed effects, controlling for district demographics (including racial composition, special education, and English learner proportions), as well as stack-by-state and stack-by-year fixed effects. All dollar amounts are adjusted to 2017 school year dollars. Standard errors are clustered at the state level, with robust standard errors in parentheses. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.)

Table 2.8: Effects on Tax and Bond Elections

	Ever had election		Pass referenda	
	Tax	Bond	Tax	Bond
Formula Cutoff	-0.002 (0.013)	-0.028** (0.012)	0.001 (0.007)	0.016** (0.007)
Proportion of Title I Kids	-0.005 (0.006)	0.007 (0.005)	0.001 (0.003)	-0.009*** (0.003)
Cutoff \times Proportion	0.003 (0.008)	-0.006 (0.008)	-0.001 (0.004)	0.009* (0.005)
N	19,813	18,175	19,813	18,175
Mean	0.336	0.403	0.201	0.262
R²	0.156	0.659	0.163	0.651

Notes: This table presents regression discontinuity estimates of Title I eligibility on tax and bond election outcomes using equation (3). The sample includes district-year observations from seven states (AR, LA, MI, MO, PA, TX, and WI) for tax elections and six states (LA, MI, MO, PA, TX, and WI) for bond elections from 2008-2017. "Ever had election" indicates whether the district held at least one referendum in the given year. "Pass referenda" represents the average passage rate of referenda if multiple elections were held in the same year, or a binary outcome of passage for districts holding a single election. All specifications include district demographics (including racial composition, special education, and English learner proportions), as well as stack-by-state and stack-by-year fixed effects. Standard errors are clustered at the state level, with robust standard errors in parentheses. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.)

Table 2.9: Effects on Tax and Bond Elections (By Title I Quartile)

	Ever had election		Pass referenda	
	Tax	Bond	Tax	Bond
Q1	0.010 (0.038)	0.020 (0.042)	0.018 (0.022)	0.030 (0.023)
Observation	3,382	2,932	3,382	2,932
Mean	0.376	0.468	0.177	0.287
Q2	-0.008 (0.024)	-0.025 (0.023)	0.009 (0.012)	0.009 (0.014)
Observation	5,259	4,840	5,259	4,840
Mean	0.321	0.450	0.196	0.298
Q3	-0.012 (0.023)	-0.022 (0.022)	-0.007 (0.012)	0.029** (0.012)
Observation	5,967	5,547	5,967	5,547
Mean	0.322	0.386	0.200	0.253
Q4	0.011 (0.023)	-0.047** (0.023)	-0.004 (0.011)	0.000 (0.013)
Observation	5,189	4,839	5,189	4,839
Mean	0.339	0.337	0.221	0.223

Notes: This table presents regression discontinuity estimates of Title I eligibility on tax and bond election outcomes by Title I funding quartile using equation (3). Districts are classified into quartiles (Q1-Q4) based on Title I amount per eligible child within each state-year using the full sample of all U.S. school districts, with Q4 representing districts receiving the highest level of Title I funding. The sample includes district-year observations from seven states (AR, LA, MI, MO, PA, TX, and WI) for tax elections and six states (LA, MI, MO, PA, TX, and WI) for bond elections from 2008-2017. "Ever had election" indicates whether the district held at least one referendum in the given year. "Pass referenda" represents the average passage rate of referenda if multiple elections were held in the same year, or a binary outcome of passage for districts holding a single election. All specifications include district demographics (including racial composition, special education, and English learner proportions), as well as stack-by-state and stack-by-year fixed effects. Standard errors are clustered at the state level, with robust standard errors in parentheses. ($*p < 0.1$, $**p < 0.05$, $***p < 0.01$.)

Table 2.10: Association Between State Funding Progressivity and Title I Effects on District Finances

	State Revenue (1)	Local Revenue (2)	Capital Outlays (3)
State Revenue Progressivity	0.035 (0.046)	0.019 (0.073)	-0.035 (0.028)
Elem/Sec Exp Progressivity	-0.010 (0.063)	0.050 (0.102)	-0.157** (0.073)
Property Tax Floor	-21.273 (42.816)	-51.288 (73.218)	80.245 (52.118)
Property Tax Ceiling	1.129 (43.249)	9.293 (74.176)	-0.299 (57.979)
Observations	49	49	49

Notes: This table presents the association between state-level funding progressivity and Title I effects on district finance. For each state, I first estimate a Title I effect by running the main regression model separately (controlling for student demographics and year-by-stack fixed effects). Next, I construct a progressivity measure by regressing per-pupil funding on a Title I indicator, weighted by Title I-eligible and non-eligible enrollment (Chingos & Blagg, 2017; Knight & Mendoza, 2019; Lee et al., 2022). This measure captures the average funding difference between Title I and non-Title I students. Finally, I regress the state-level Title I effect on the progressivity measure, weighting each state by the inverse of the squared standard error of its estimated Title I effect. Standard errors in parentheses. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.)

Chapter 3

PAPER TWO: DOES COREQUISITE REMEDIATION WORK FOR EVERYONE? AN EXPLORATION OF HETEROGENEOUS EFFECTS AND MECHANISMS (WITH DR. FLORENCE XIAOTAO RAN)

Abstract The landscape of developmental education has experienced significant shifts over the last decade nationwide, as more than 20 states and higher education systems have transitioned from the traditional prerequisite model to corequisite remediation. Drawing on administrative data from Tennessee community colleges from 2010 to 2020, this study examined the heterogeneous effects of corequisite reform for remediation-eligible students with varying levels of academic preparation. Using difference-in-differences and event study designs, we found that corequisite remediation significantly improved gateway and subsequent college-level course completion for students in all placement test score groups below the college-level threshold. For math, the positive effects on college-level course completion were stronger for higher-scoring remedial students than for those with lower placement test scores, whereas the pattern was reversed for English. However, since the corequisite reform, students requiring remediation were more likely to drop out of the public college system, and those with the lowest scores were less likely to earn short-term certificates.

3.1 Introduction

Academic under-preparation is one of the main non-financial barriers to college access and success (Dynarski et al., 2022). Of the 2019 graduating class—the last high school cohort who did not experience disruptions of the Covid pandemic—an estimated one in every three graduates did not meet the college-readiness benchmarks in any subject tested (ACT, 2019). Because of the open-door policy of community colleges and other broad-access postsecondary institutions, developmental education (or remediation), a program designed to bring underprepared students up to an adequate level for college study, should be instrumental in ensuring mass postsecondary education. Since Black and Hispanic students and students from low-income backgrounds disproportionately enroll in developmental courses (Chen et al., 2020), this program should also play an important role in mitigating racial and income gaps in college enrollment and completion.

However, the landscape of developmental education has experienced significant shifts across the nation in the last decade, mainly because the traditional remediation model did not achieve its intended goals. The traditional prerequisite approach typically required students below college-ready thresholds to pass a sequence of remedial courses before enrolling in college-level coursework. However, this approach had a few problems, including inaccurate placement (Scott-Clayton et al., 2014), high attrition rates (Bailey et al., 2010), and disconnected content covered in remedial and college-level courses (Cullinane & Treisman, 2010). In the past few years, more than 20 states have adopted corequisite models, allowing remedial students—and in many cases, mandating everyone regardless of their level of readiness—to take college-level courses with concurrent academic support upon initial enrollment (Education Commission of the States, 2021). Such reforms led to significant improvements in the completion rates of the first college-level courses, especially for students who just missed college thresholds by a few points (Meiselman

& Schudde, 2022; Ran & Lin, 2022).

The gap in understanding of corequisite reform is its implications for students who are far below the college-readiness threshold. These students were typically excluded from previous literature examining the effects of developmental education, as most previous research relied on regression discontinuity (RD) designs using students just above and below the college-ready threshold to establish causality. Two studies did extend their analyses to lower-scoring students, but what they examined was the “local” effects of a shorter vs. a longer prerequisite sequence, not the effects of mainstreaming all remedial students into college-level classrooms (Boatman & Long, 2018; Xu & Dadgar, 2018). Despite a few randomized control trials showing overall positive impacts (Logue et al., 2019, 2016; Miller et al., 2022), researchers and higher education practitioners need more causal evidence of the heterogeneous impacts of removing standalone prerequisite sequences.

This study closes this gap in the literature by examining the heterogeneous effects of corequisite versus prerequisite remediation on college success and unpacking the mechanisms of how corequisite remediation affects students with different levels of college readiness. To do so, we obtained administrative data from the Tennessee Board of Regents (TBR) between 2010-11 and 2019-20. TBR provides a great context for this study, as it was the first higher education system in the nation to replace the standalone prerequisite sequence with the corequisite models for all incoming students since 2015. We used difference-in-differences (DID) and event study strategies to compare the outcomes of remedial students with similar placement scores before and after the corequisite implementation, using changes in the outcomes of college-ready students during the same time period to control for general time trends or shocks from other policy changes. To account for slight variations in corequisite implementation timelines across colleges, we also used heterogeneity-robust estimators for our DID and event study models as robustness checks (Callaway & Sant’Anna, 2021; Sun & Abraham, 2021)

Overall, we found that students placed into corequisite remediation were up to 20.4 percentage points (or 77%) more likely to pass gateway math and 22.8 percentage points (or 42%) more likely to pass gateway English within one year of enrollment compared with otherwise similar students placed into prerequisite remedial courses. However, corequisite reform had null effects on the number of college-level credits earned or the likelihood of transfer to a public four-year university by the end of the third year. Since the corequisite reform, students placed below the college-readiness threshold were 4.3 percentage points (or 8.1%) less likely to continue enrolling in the state’s public college system and 3.0 percentage points (or 28.8%) less likely to earn a credential (mostly certificates) within three years of the initial enrollment. While the effects on gateway course completion were positive across the full spectrum of placement test score distributions, the negative effects on enrollment persistence and credential completion were primarily driven by students coming in with the lowest scores.

We make three distinctive contributions to developmental education research and policymaking. First, we provided new empirical evidence on the effects of the developmental education reform for students who were not present in the previous literature, primarily using RD methods. As shown in Table B.1, the analytic samples in the previous literature include students within a few points above and below the college-readiness threshold, typically between the 40th and 50th percentiles, if converted to ACT. In contrast, our study included students on the entire spectrum of placement score distribution, with ACT composite scores ranging from 1 to 34. By including these students, our sample covers a higher proportion of underrepresented minorities, which represents approximately 55% of the lowest scoring group. Researchers have repeatedly shown that students at the lower end of standardized test score distributions are more likely to come from low-income backgrounds or live in poverty (Dixon-Román et al., 2013). Evidence of the effects of the corequisite reform for students in these demographic groups is crucial to understanding the potential

of one of the most extensive interventions targeting academic under-preparation in closing racial and socioeconomic gaps in college success.

Second, we articulated a conceptual framework of how corequisite remediation leads to changes in college outcomes, built on theoretical frameworks developed by Scott-Clayton and Rodriguez (2015) and Kane et al. (2019). We found that the positive effects on gateway course completion were driven by providing direct access and reducing delays to college-level courses. Since students in the lowest-scoring group used to go through the longest prerequisite sequence, they experienced the largest improvement in gateway course enrollment among all remedial students. However, not all students were able to pass the gateway and more advanced college-level courses. The total instruction time on developmental content significantly reduced since the reform, almost cut in half for students with lower scores. In addition, corequisite models also produced de-tracking effects, making college-level classrooms more heterogeneous in terms of students' academic readiness. We found that a 10-percentage-point increase in on-level peers was associated with a 2.7-percentage-point decrease in gateway math completion rates and a 3.3-percentage-point decrease in English.

Lastly, these results can inform ongoing reforms across more than twenty states and systems. The heterogeneous effects that corequisite reform produced for students from different test score groups as well as for different subjects suggest that there is no one-size-fits-all corequisite model. For students who are severely underprepared for college-level studies, colleges need to consider ways to provide enhanced support such as tutoring and academic counseling. This support may need to continue after students move beyond their first college-level math and English, as our results indicated that many lower-scoring remedial students were not able to pass advanced college-level courses even after they successfully completed gateway courses. The more sequential nature of math skill development compared with English means that colleges may experiment with alternative structures of

corequisite learning support for different subjects. Some colleges and systems have been adopting corequisite models that use the first half of the semester to cover fundamental skills and move to college-level materials in the second half. Colleges may also consider designating certain gateway course sections exclusively for students with a greater need for support, and adapting the course pace accordingly.

3.2 Literature Review and Conceptual Framework

3.2.1 Previous literature on developmental education

Traditional prerequisite developmental education (DE) is essentially a tracking system that affects student outcomes in three ways: skill development that prepares students with remedial needs for future college-level courses, a delay that postpones enrollment in college-level courses, and a diversion that places remedial and college-ready students into separate courses and reduces heterogeneity within classrooms (Kane et al., 2019; Scott-Clayton & Rodriguez, 2015). Ideally, students should be placed in the appropriate track to develop skills that satisfy their needs. The positive developmental boost has to be large enough to offset the delay or diversion from college-level coursework for developmental education to provide an overall positive effect on college outcomes.

As shown in Panel A of Table B.1, the majority of prior literature found null to negative effects of prerequisite DE on early college outcomes. The prerequisite nature of the support created multiple exit points before many students getting into the college-level study: as Bailey et al. (2010) estimated, up to one third of students referred to DE exited the prerequisite sequence before completing the sequence. Many of them could have passed college-level courses without any remediation, as they were misplaced into DE by a single standardized test score (Scott-Clayton et al., 2014). These students suffered from the adverse consequences of delay, with little or no developmental benefit. Recent randomized trial evidence found that compared to single measures alone, multiple measures that combined high school

GPA, standardized test scores, and other measures for placement purposes were much more predictive of future performance (Barnett et al., 2020; Bergman et al., 2021).

Another problem with the traditional DE was its instructional approaches and curriculum designs. Many community college students did not understand basic algebra concepts or have the literacy skills to read college-level textbooks. Traditional DE courses did not provide enough developmental benefits for students truly in need of academic support to succeed in college-level courses because of decontextualized instruction and poor alignment between remedial and college-level courses (Jaggars & Bickerstaff, 2018). To improve college outcomes, remedial students also need a more effective curriculum and pedagogy to garner skill development.

The emerging evidence in general showed that corequisite approaches were more effective than the traditional DE in helping students pass the first college-level English and math courses (Panel B of Table B.1). Except for two randomized control trials with smaller sample sizes (Logue et al., 2019; Miller et al., 2022), studies on corequisite DE have focused on the around-the-cutoff population. To inform the current waves of corequisite DE reforms across the nation, researchers and practitioners need to better understand the following areas: (1) the effects of corequisite DE for students at the lower end of skill distribution, (2) the ways in which corequisite DE affects the curriculum and total instructional time for students with different levels of preparedness, and (3) whether corequisite DE works similarly for different subject areas.

3.2.2 Conceptual framework of the effects of corequisite DE

Corequisite DE leads to changes in all the three mechanisms described above. First and foremost a structural reform, corequisite remediation is premised on reducing the delay effect by allowing all incoming students to enroll in college-level study immediately upon enrollment. This would also lead to more students enrolling

in college-level courses, as this approach helps students avoid exiting points before college-level courses. In addition, with the corequisite approach, college-ready and remedial-eligible students are mixed in college-level classrooms, which changes peer dynamics and may result in differences in learning outcomes for both on-level and remedial students after the corequisite reform.

As for the “skill development” mechanism, we hypothesized that corequisite DE leads to changes in the timing, content, and intensity of developmental support. Corequisite DE provides “just in time” support to remedial students, rather than asking them to build the required skills in a prerequisite sequence before they arrive at college-level classrooms. Due to the contemporaneous nature of the support, many colleges use corequisite reform as an opportunity to align the content and curriculum of developmental support with college-level courses. One example is that incorporating alternative math pathways and content in developmental math no longer focuses solely on algebra. Colleges offer corequisite learning support for statistics, quantitative reasoning, and math for liberal arts to pair with the gateway math courses required by students’ program of study. Finally, the corequisite approach typically compresses the total instructional time for remediation by combining two or more developmental courses into a single one-semester experience (Edgecombe et al., 2013).

Both the structure and quality of implementation would determine how the skill development mechanism works for corequisite remediation. Originating from Accelerated Learning Programs for English (Cho et al., 2012), the premise of corequisite learning is that contemporaneous support can help students build and refresh skills in meaningful contexts before they fade out. However, curricular and skill development may be more sequential in certain subjects than in others. Students with lower placement test scores arrive at community colleges lacking proficiency in fractions, decimals, and applications of algebra skills in word problem solving (Ngo, 2019). When these students are enrolled in a college algebra or statistics course,

just-in-time support may not guarantee work.

In addition, while this study primarily focuses on the structural perspective of the corequisite reform, we must not overlook the crucial role of instructors in teaching learning support courses and the pedagogical practices employed in these courses, as they heavily influence the extent to which corequisite remediation benefits students' skill development. One challenge in replacing standalone DE with corequisites is the shortage of faculty with credentials to teach college-level coursework, especially if colleges want to staff a common instructor for the corequisite and college-level sections (Daugherty et al., 2018). This problem may be more pronounced for math than for English, as it is harder for STEM fields to recruit high-quality instructors when earnings and job opportunities in alternative industries are more competitive for individuals with STEM degrees (Xu & Ran, 2022). These factors may contribute to the different effects of corequisite DE across different subject areas.

These three mechanisms may function differently for students with different levels of preparedness. Students with lower placement test scores experienced the longest delay from the traditional DE, as they were assigned to longer DE sequences. Corequisite DE may benefit them the most in terms of college-level credit accumulation, by giving them direct access to these courses. The effects of mixing remedial and on-level students in college-level classrooms depend on the peer composition of course sections. Previous literature provides a strong theoretical and empirical basis for the existence of peer effects in higher education (Sacerdote, 2011; Winston & Zimmerman, 2004), but it is unclear whether underprepared students could benefit more from higher-achieving students because of the transfer of specific knowledge and general academic know-how (Griffith & Rask, 2014) or from similar-ability students as instructors could teach to their level. The development component under corequisite models may work better for higher-scoring remedial students, as previous studies have shown that the lowest-placed remedial students benefit from an intensive focus on building basic numeracy and literacy skills before matriculating

(Scrivener et al., 2018).

3.3 Context & Data

3.3.1 State & institutional context

Remediation placement. The placement policy for developmental education in Tennessee did not undergo any major changes during the study period. Scores from the mandatory standardized tests in 11th grade, mainly ACT, were the sole determinant for remediation at TBR community colleges until the fall of 2020. The ACT scores to satisfy direct placement in college-level remained as a score of 19 for math, 18 for writing, and 19 for reading between 2010 fall to 2020 spring. Since the majority of TBR community colleges set minimum scores for both writing and reading for direct placement into college-level English, we used the minimum scores of the two subjects to determine students' placement status for English. It is possible for students to re-take the ACT during high school or take other types of standardized tests (such as ACCUPLACER, ASSET, or COMPASS) after arriving at college to waive remediation requirements. In our data, around 6% of the entering cohorts between 2010 and 2018 reported multiple test scores. As students who challenged their remediation status by retesting might be systematically different from others, we used students' earliest available scores on records for the analyses.

Corequisite implementation. During Fall 2014 and Spring 2015, the TBR started to pilot corequisite remediation across nine colleges (Denley, 2015). We found that 11 out of the 13 community colleges fully scaled corequisite implementation in academic year 2015-16, with virtually no students taking standalone prerequisite remedial courses afterward, while the remaining two colleges did so by 2017-18. Before the corequisite reform, remedial students were placed into a developmental sequence consisting of different levels and numbers of courses based on their placement test scores. Our analyses of student transcript data suggest that the number of prerequisite remedial credits enrolled ranged from one course for those just below the

college threshold, to up to three courses for those with the lowest ACT scores. After the reform, while TBR's policy allowed colleges to offer varying levels of corequisite learning support, the system's internal report showed that colleges rarely tailored the learning support experience based on a student's academic preparation. Most corequisite courses are semester-long experiences for all remedial students. TBR does not require a standard format for learning support courses. We found some variations in the delivery methods, with three colleges offering the vast majority of their corequisite courses through online courses. We did not see systematic changes in corequisite implementation in TBR's policy documents until 2020 Fall, when the system started a pilot to add high school GPA as a method for placement alongside traditional measures such as ACT scores.

Other policies aimed at improving college success. During the same period of the corequisite implementation, the public education system in Tennessee enacted a few other policy changes to improve college access and success. First, public high schools in Tennessee launched the Seamless Alignment and Integrated Learning Support (SAILS) program in 2012, which aimed to help students complete math remediation in high school and was scaled to around 250 high schools in 2015. Kane et al. (2021) showed that completing SAILS boosted enrollment in college-level math for the college entering cohort of 2014-15, but after the corequisite policy took effect, SAILS no longer produced any additional effects. To separate the possible influence of SAILS, we conducted robustness checks using a subsample of students who graduated from high schools and were unlikely to be affected by SAILS. The Tennessee Promise Scholarship was launched in 2015. It provides Tennessee high school graduates attending community colleges full-time with last-dollar scholarships to cover tuition and fees. As discussed in more detail in the Empirical Strategy section below, this would not bias our results as long as the Promise Program affects students above and below the college readiness threshold in a similar way.

3.3.2 Data Description

Table B.2 presents the sample restriction process and how summary statistics change across different samples. Our raw data came from 407,193 students enrolled at one of the 13 community colleges in TBR at any time since the fall of 2010. Given our focus on the heterogeneous effects of corequisite remediation across test score distributions, we restricted our analyses to first-time-in-college students who had at least one set of standardized test scores on record. We also excluded students who entered TBR later than the fall of 2018 to allow a sufficient tracking period for enrollment and attainment outcomes. This restriction dropped around half of the students in the raw data, leaving us a sample of 194,524 students who were much younger at college entry (19.5-year-old vs. 24.0-year old). Next, we restricted our sample to first-time-in-college (FTIC, hereafter) students who entered TBR community colleges during the fall semester immediately after high school. We implemented this restriction for two reasons. First, this is to mitigate possible changes in students' age composition across cohorts due to the introduction of the Promise Scholarship in 2015, as students must apply for Promise during their senior year in high school. In addition, the recent high school graduate fall entrant sample was significantly less prone to data quality issues. Only 2% of the students in this sample did not report a 4.0-scale high school GPA (because of missing or non-standard format high school GPA), whereas 11% of the FTIC sample and close to half of the students in the raw data did not have valid high school GPA records. The caveat of these sample restrictions is that compared to the general student population that TBR community colleges serve, students in our final analytic sample were much younger and had slightly lower test scores and high school GPAs. We need to be cautious about generalizing the results of this study to older students who have longer gaps between high school and community college enrollment.

[Table 3.1 about here]

The final analytic sample consists of 91,511 students. Table 3.1 contains descriptive statistics for all background and outcome variables for students in our analytic sample by prerequisite and corequisite cohorts. In general, students' demographic characteristics remained fairly consistent across cohorts. On average, 75% of students were assigned to remediation for at least one subject, and the proportions of students referred to remediation remained consistent for students entering TBR during prerequisite and corequisite regimes. On the other hand, changes in outcome variables between the prerequisite and the corequisite cohorts are noticeable. Corequisite cohorts were more likely to complete gateway courses by the end of the first year, by 14 percentage points for math and 9 percentage points for English. In addition, corequisite cohorts were more likely to enroll and complete additional college-level math or English courses. However, the differences in college-level credit accumulation between the corequisite and the prerequisite cohorts diminished by the end of year three. No clear patterns emerged for persistence and credential completion outcomes: the likelihood of stopping enrollment at community colleges and public universities in Tennessee increased by 10 percentage points for the corequisite cohorts, but credential completion rates also increased by 5 percentage points.

To assess the heterogeneous effects of corequisite versus prerequisite remediation by readiness, we separated students below the college-readiness threshold into three subgroups based on placement test scores converted to ACT scales. Group 1 contained students with the lowest scores, who scored six or more points below the college readiness threshold (15% of the analytic sample). Group 2 included students scoring three to five points below the cutoff (38% of the analytic sample), and Group 3 was the highest-scoring remedial students who scored one or two points below the college-readiness cutoff (23% of the analytic sample). The rationale for test score grouping is as follows. A previous study on corequisite reform in Tennessee examined the effects of corequisite versus no remediation for Group 3 students using the regression discontinuity (RD) method (Ran & Lin, 2022). These students were

virtually identical to students right above college-level thresholds in terms of demographic and high school academic measures. Groups 1 and 2 contained students with weaker academic preparation, as reflected by their lower placement test scores and lower average high-school GPA. We chose the six-points- from-cutoff threshold to separate Groups 1 and 2 because most students (j , 98%) below this threshold were referred to corequisite learning support for all three subjects. It takes 18 credit hours to complete three gateway courses paired with learning support sections. Since most full- time students register for four courses (or 12 credit hours) in one semester, it is very difficult for students in Group 1 to fulfill the corequisite requirements within the first semester.

3.4 Empirical Strategy

3.4.1 Identification Strategy

We applied a difference-in-differences (DID) approach to estimate the impacts of corequisite remediation on a set of college success outcomes, compared with traditional prerequisite remediation. Exploiting that any changes in remediation policy should affect only students below college level, we compared the outcomes of remedial-eligible students before and after the implementation of the corequisite reform, using college-ready students as the control group to adjust for any general time trends or policy changes (such as Promise Scholarship) that potentially affected all students. Specifically, we estimated the following equation:

$$Y_{ijt} = \beta_0 + \beta_1(Below_i \cdot Post_{tj}) + \beta_2Below_i + \phi_t + \lambda_j + \beta_n\mathbf{X}_{itj} + \varepsilon_{itj} \quad (3.1)$$

Here, Y_{ijt} is the outcome of student i of entering cohort t in college j . $Below_i$ is a student-level indicator for placing below college-level cutoff scores. For the student-level analyses, such as enrollment persistence or credential completion, we used minimum scores across all three subjects to define placement into remediation.

$Post_{tj}$ is an indicator for post-corequisite cohorts, which varies across colleges. The coefficient β_1 captures the intent-to-treat (ITT) estimates of the effects of placement into corequisite remediation versus the prerequisite model. In this model, we also controlled for any general time trends through cohort fixed effects (ϕ_t) and any systematic differences in student outcomes across institutions through college fixed effects (λ_j). The vector \mathbf{X}_{itj} contains a set of student covariates including gender, age at college entry, race, international student status, placement test scores, high school GPA, and high school fixed effects. Standard errors were three-way clustered at college-, high school-, and cohort- level.

To capture the heterogeneous effects of corequisite remediation by academic preparedness, we further estimated the following model:

$$Y_{itj} = \gamma_0 + \gamma_k(\text{Scorebin}_i \cdot \text{Post}_{tj}) + \gamma_n(\text{Scorebin}_i) + \phi_t + \lambda_j + \beta_n \mathbf{X}_{itj} + \varepsilon_{itj} \quad (3.2)$$

where Scorebin_i represents the indicators for the three groups of placement test score bins described above, and the coefficient γ_k captures the effects of placing into corequisite versus prerequisite remediation for students within each bin. The reference group consisted of students above the college-ready threshold.

Because not all remedial-eligible students enrolled in the remedial courses they were referred to, we can obtain the enrollment effects of corequisite versus prerequisite remediation (i.e., treatment on the treated) by instrumenting a student's enrollment in remedial courses using their placement status. The classic exclusion restriction assumption in the instrumental variable framework requires that a student's remedial placement status does not affect her outcomes in ways other than enrollment in remedial courses. Since we can only observe students already enrolled in community colleges in our sample, it requires that remedial designation did not have a direct impact on students' decisions to attend college. Martorell et al. (2015) provided evidence that the effects on college enrollment were insignificant

for students around the college-level threshold, although it is not clear whether such conclusions can be generalized to students further away from the college-level threshold. Therefore, in this paper, we focus on the reduced-form estimates of placement effects as our main results.

We also used event study models to examine the effects of corequisite remediation over time. The regression model used for the analysis is as follows:

$$Y_{itj} = \phi_t + \lambda_j + \sum_{\tau=-4}^{-1} \gamma_{\tau} \text{Below}_{itj} + \sum_{\delta=0}^3 \gamma_{\delta} \text{Below}_{itj} + \beta_n \mathbf{X}_{itj} + \varepsilon_{itj} \quad (3.3)$$

The coefficient γ_{τ} and γ_{δ} measures the differences in trajectories of outcomes between remedial and college-ready students relative to the corequisite implementation timeline after controlling for time fixed-effects (ϕ_t) and college fixed-effects (λ_j). Across all specifications, we used five years before the corequisite implementation as the reference year to estimate γ_{τ} and γ_{δ} .

3.4.2 Validity of DID estimates

Our setting includes college-ready students who were never treated by corequisite remediation and remedial students treated by the reform that started at different time points across colleges. As Goodman-Bacon (2021) showed, the traditional two-way fixed effects (TWFE) estimator described in Equation (3.1) is a weighted average of all potential canonical 2x2 DID estimates where weights are based on group sizes and variance in treatment. With a staggered treatment timing setting, the following assumptions were required:

The first is the variance-weighted common trends. This is similar to the parallel trend assumption in the canonical DID settings. In other words, the trajectories of outcomes for remedial and college-level students would have been the same if corequisite remediation had not been implemented. We examined whether remedial students were already on a different trajectory before the implementation

of corequisite remediation by testing whether γ_τ in Equation (3.3) was statistically significant. The results in Figure B.1 suggest that, compared with five years before the corequisite reform, the effects of placing below the college-readiness threshold were insignificant for two to four years before the reform, and as expected, gateway completion rates started to increase one year before full implementation during the pilot period. These event study estimates provided evidence supporting the parallel trends assumption, as they suggested that the outcomes of remedial and college-ready students had similar trajectories before the reform.

To isolate the effects of corequisite remediation, another important assumption is that the characteristics of students in the treatment and comparison groups did not undergo differential changes during the study period. To test this, we conducted a series of covariate balance tests using our main difference-in-differences specifications with student covariates as outcomes. In these tests, the coefficients of below college-readiness cutoff tell us about the overall differences in covariates between college-ready and remedial students. Meanwhile, the coefficients of post corequisite reform show how student characteristics change over time in general. The coefficients of the interaction terms between below the cutoff and post-reform are the parameters of interest here: they show whether student characteristics were significantly different for students below the college readiness cutoff after the reform. The results reported in Table B.3 suggest that changes in all but one covariate (the proportion of black students) were balanced across the treatment and control groups. We further conducted subgroup covariate balance tests to examine whether changes in student characteristics systematically differed across remedial students in the different test score groups. The results in Panel B of Table B.3 show that the lowest test score group contained fewer underrepresented minority students than before the reform. In addition to the small changes in racial composition, the three groups of remedial students and college-ready students did not experience systematic differences in student demographic characteristics and academic attributes.

This assumption also requires that the corequisite reform or other policies during the same period, such as SAILS, did not affect college-going decisions or alternate characteristics of college enrollees over time. While we cannot directly conduct a test on this, as our data only contain students already enrolled in TBR colleges, Kane and coauthors provided evidence that SAILS did not affect enrollment in community colleges by the spring of 1st year after high school graduation (Table 3 of Kane et al. (2019)). Taken together, compositional changes in the comparison group are unlikely to be an alternative explanation for the findings presented below.

The next assumption is the constant treatment effects within groups over time. As discussed in Goodman-Bacon (2018), TWFE estimates could yield biased results when the composition of the comparison group is changing (representing a shifting mix of not-yet- implementers and previous-implementers) if the treatment effects change over time within groups. To assess whether the effects of the corequisite reform changed over time, we conducted F-tests to examine whether the coefficients of placing below the college-readiness threshold were equal across the corequisite implementation timeline. The F-statistics were 1.51 for the completion of gateway math by year one and 0.88 for English. We cannot reject the null hypothesis of homogeneous effects over time. Nevertheless, we further examined whether our results were robust in the case of varying effects between early and late adopters. We compared the estimates from the original TWFE models with two different heterogeneous-robust estimators recently developed by Callaway and Sant’Anna (2021) and (Sun & Abraham, 2021), as shown in Columns 3 and 4 of Table 3.2. Figure B.1 illustrates comparisons of these estimators. The fairly consistent results across the models suggest that the TWFE specifications are robust in our setting.

3.5 Results

3.5.1 Effects of corequisite remediation

Gateway course completion. Table 3.2 shows the effects of corequisite versus prerequisite remediation on first-year gateway math and English completion rates for all students in our analytic sample. Based on the TWFE estimates, students placed into corequisite remediation were 20.4 percentage points more likely to complete gateway math by the end of year one, compared with otherwise similar students placed into prerequisite remediation. This represents a 76% improvement from the baseline average of the prerequisite cohorts. The effect for English was 22.8 percentage points (41% increase from the baseline mean). The event study estimates (Table 3.2 Panel B) show how the effects of placement into remediation changed relative to the corequisite implementation timeline, using five years before the reform as the reference group. As discussed in the previous section, we take the insignificant coefficients for two to four years before the corequisite pilot and reform as evidence for the parallel trends assumption, as they suggested that there were few changes in remedial students' outcomes before the reform. The heterogeneity-robust estimates reported in Columns 3 and 4 in Table 3.2 are consistent with the TWFE estimates. Because the computational package producing TWFE estimates allows for more flexible covariate controls and multi-way clustering of standard errors, we hereafter use TWFE specifications as our preferred model.

[Table 3.2 about here]

One of our main goals is to examine the effects by student preparedness; in Table 3.3, we report the heterogeneous DID estimates by placement scores. As described in the Context and Data section, Group 1 contained students scoring six or more points below the cutoff, Group 2 included students scoring three to five points below the cutoff, and Group 3 was the highest-scoring remedial students, within two points below the college-ready threshold. These results suggest that, compared with

prerequisite remediation, corequisite remediation had significant positive effects on first-year gateway completion rates for remedial students in all test score groups below the college-ready threshold. For math, the effect was strongest for the highest-scoring remedial students. Students who scored within two points of the college-level cutoff experienced a 21-percentage-point improvement in the first-year gateway completion rate, while the lowest-scoring remedial students had a 15-percentage-point improvement in the first-year gateway completion rate. For English, the pattern was reversed. The lowest-scoring students experienced the greatest improvement in first-year gateway English completion rate (up to 33 percentage points).

[Table 3.3 about here]

Subsequent college-level courses. Next, we examined the effects of corequisite versus prerequisite remediation on subsequent college-level course enrollment and performance. Table 3.4 presents the results. For both math and English, students in the corequisite cohorts were significantly more likely to enroll in and pass a second college-level course in math and English by the end of year two. Interestingly, the grades in subsequent college-level courses were significantly lower after the corequisite implementation. This does not necessarily mean that corequisite remediation had negative effects on course performance in subsequent college-level courses, because these analyses were conditional on students enrolling in a second college-level course. During the prerequisite era, students who managed to do that—note that they had to fulfill remedial sequence requirements and pass the gateway course before that—were likely to be systematically different from students who got to second college-level courses under the corequisite policy. However, this result highlights the challenge of implementing corequisite remediation. Since every student can enroll in college-level courses under corequisite models, community colleges must either increase course offerings or increase enrollment caps for college-level course sections.

In the TBR, community colleges kept enrollment sizes in college-level courses stable over time, but the number of course sections for college-level math increased by more than 30% between the fall of 2014 and the fall of 2015. How to schedule the additional course sections, staff faculty with credentials to teach them, and provide quality instruction to all students are key issues.

[Table 3.4 about here]

Again, the subgroup analyses (Table 3.4, Panel B) showed reversed patterns for math and English. The positive effects on enrollment in and completion of subsequent college-level math courses were mainly driven by higher-scoring remedial students. Students who scored within two points of the college-level cut-off were six percentage points (or 52%) more likely to enroll in and five percentage points (or 43%) more likely to complete a second college-level math course during the corequisite era than the prerequisite cohorts, while the effects for the lowest-scoring remedial math students were insignificant. For English, it was the lowest-scoring group that experienced the strongest improvement, by 15 percentage points (93%) for enrollment and 9 percentage points (73%) for passing a second college-level English course.

Longer-term outcomes. To assess the effects of corequisite remediation on long-term outcomes, we conducted analyses on a set of outcomes related to credit accumulation, enrollment, and completion tracked over three years after initial enrollment. For credit accumulation, corequisite remediation led to 1.5 fewer total credit enrollments, but 2.5 more college-level credits enrolled; the effects on total college-level credits earned were insignificant (Columns 1 to 4 of Table 3.5). As for persistence and completion outcomes, the general effects of corequisite remediation tended to be negative: remedial students under the corequisite regime were 4.3 percentage points more likely to drop out of college, defined as stopping enrolling in any TBR community college or public four-year university before earning a credential,

at the end of year three. They were also three percentage points less likely to complete any credential by that time (Columns 5 to 7 of Table 3.5). The heterogeneous analyses (Table 3.5 Panel B) suggest that the effects on the number and composition of credit accumulation came from a reduction in the total number of credits attained, mostly driven by enrolling and earning fewer developmental courses, for the highest-scoring remedial students (2.1 credits or close to one course) and more college-level credits enrolled for the lowest-scoring remedial students (5.1 credits or close to two courses). The negative effects on enrollment at community colleges and public universities in Tennessee were consistent across all test score groups, but the negative effects on credential completion mostly came from lower-scoring remedial students. This was primarily driven by the lower likelihood of earning short-term certificates: more than 70% of the credentials earned by the lowest-scoring students within three years were short-term certificates, and the proportion of students earning degrees at the associate level or higher did not change significantly before and after the corequisite reform.

[Table 3.5 about here]

Robustness checks. We conducted the following analyses to assess the robustness of the main results. First, we ran analyses excluding the data from the academic year 2019-20 to rule out the possible influence of the Covid-19 pandemic. The results showing the effects on gateway course completion and long-term outcomes are presented in Tables B.4 and B.5, respectively, and are consistent with the main results discussed above. In addition, since the SAILS program serves a purpose similar to corequisite reform, we conducted a robustness check excluding students attending high schools where SAILS was introduced during their senior year. Since SAILS eventually reached most public high schools within the state, this analysis excluded more than half of the students who entered TBR after the academic year 2016-17.

As shown in Table B.6, the results for gateway courses, enrollment persistence, and credential attainment are similar between models with and without high schools that have introduced the SAILS program. The most noticeable disparities between the two sets of analyses are for outcomes related to credit accumulation by the third year. For the sample that excludes students from high schools with a heavy SAILS presence, the effects of the corequisite policy on the total number of credits attained were insignificant, and the effects on college-level credits enrolled and earned were significantly positive. The discrepancies are likely due to the corequisite reform muting any boost in college credits resulting from completing SAILS, aligning with the findings of Kane et al. (2021).

3.5.2 Implications of corequisite reform effects

The effects of corequisite remediation on gateway course completion and long-term outcomes can be interpreted from two perspectives: (1) Does corequisite remediation help students finish courses count towards a degree, and (2) is corequisite remediation—or developmental reforms in general—enough to move the needle to improve college completion? From the first perspective, the answer appears to be yes. Under the corequisite models, students were able to earn similar amounts of college-level credits while enrolling in fewer courses. Credits from developmental courses cannot be applied to a degree; in themselves, these courses, especially developmental math, even have negative labor market returns (Hodara & Xu, 2016). Overall, spending valuable time and financial aid resources at college-level rather than remedial courses was a more efficient way to allocate resources. This is perhaps why a metric called “throughput,” defined as the completion rate of college- or transfer-level English and math among a cohort, is the primary yardstick to measure the success of DE legislation, such as AB705 in California and Senate Bill 1720 in Florida (Melguizo et al., 2021; Zhao et al., 2022).

However, our results suggest that colleges need to think beyond what throughput rates measure and place DE reforms in the context of the broader sets of institutional support to help students complete a postsecondary degree, as replacing prerequisite remediation with corequisite models alone was not enough to solve the completion problem. Mainstreaming every student in a college-level course presents new challenges. The average grades in college-level math (including both gateway and more advanced courses) for remedial students went from 1.91 in the prerequisite era to 1.50 for the corequisite cohorts, and the proportion of those who failed any of these courses increased from 36.9% to 50.2%. Early course failures often lead to discouragement or disinterest. Failing college-level courses may also have financial consequences. Many need-based financial aid programs, including Pell Grants and the Tennessee Promise, require students to maintain a certain level of GPA at the end of each academic year. In this new context, institutions need to provide more proactive advising, academic counseling, and other support to students who are unable to meet the expectations of more challenging college-level courses.

3.5.3 Exploration of effects mechanism

In this section, we examine the potential mechanisms through which the corequisite reform worked. In Table 3.6, we show the timing of remedial students enrolled in their first gateway courses before and after the corequisite reform by placement test score groups. For math, it was common to start enrolling in gateway courses until the third or fourth term during the prerequisite regime, and most remedial students (more than 70% of the lowest-scoring group) did not enroll in gateway math courses during the tracking period. After the corequisite reform, the majority of remedial students enrolled in gateway math, and most students managed to enroll during the first two terms. This reduction in delay was even more substantial in English. While the proportion of students enrolled in gateway English in their first two terms ranged from 28.5% for the lowest-scoring group to 72.2% for

the highest-scoring group in the prerequisite era, close to 80% of corequisite students enrolled in gateway English in the first term.

[Table 3.6 about here]

We then examined other factors contributing to how corequisite remediation helps students develop skills, including how the intensity and content of remedial courses changed through corequisite models. The results in Table 3.7 suggest that the total instructional time for developmental education has decreased since the corequisite reform, especially for those with lower placement scores. Since the reform, the total number of remedial credits enrolled decreased from 13.6 to 7.2 by the end of the third year for the lowest scoring group. It would be a challenge for instructors to cover the content and skills needed to be successful in college-level studies when the instruction time almost halved. This could contribute to the patterns in grades of gateway courses and more advanced college-level courses after the corequisite reform.

[Table 3.7 about here]

In addition to the compression of the development course sequence, the developmental math curriculum also underwent major changes. Before the corequisite reform, the developmental math sequence typically started with a Foundation of Math or Basic math for the lowest-scoring students, an Algebra I course for moderately scoring students, and an Intermediate Algebra course for the highest-scoring remedial students. After the reform, the content of the learning support courses was aligned with that of math gateway courses. Consequently, Corequisite Statistics has become the most popular math learning support course since 2015. Across all test score groups, more than 40% of the remedial students took this course as a companion of gateway math on the statistics pathway. Based on the archive of the course catalog, prerequisite developmental math courses designed for lowest-scoring students covered topics including addition, subtraction, division, multiplication, metric

notation, factions, and formulation solving. Our results on remedial students' performance in college-level courses suggest that when students needed refresh of these basic skills, contemporaneous support was insufficient to improve downstream college outcomes.

[Figure 3.1 about here]

Next, we examined how de-tracking influenced the outcomes for remedial and college-level students, as corequisite remediation mixes students with all levels of readiness together in college-level courses. Figure 3.1 presents the proportion of on-level students who were above the college threshold or had completed remedial requirements in gateway course sections by academic year. Indeed, since 2015, the proportion of on-level students has decreased significantly in both math and English gateway course sections. Table 3.8 shows the associations between the proportion of on-level peers and the first-year gateway course completion rates. These results have two main implications. First, students experienced better gateway course outcomes when their sections had more high-ability (on level) students. Overall, a 10-percentage-point increase in on-level peers in the section was associated with a 3-percentage-point increase in gateway math completion rates and a 2-percentage-point increase in gateway English. Second, perhaps because gateway classrooms became more heterogeneous after the corequisite reform, remedial students had higher gateway completion rates when they studied with more similar-ability (remedial) students since the reform. For students placed into corequisite remediation, a 10-percentage-point increase in on-level peers was associated with a 2.7-percentage-point decrease in gateway math completion rates. The patterns were similar for English, except that the results were entirely driven by remedial students, as the outcomes for college-ready students in gateway English courses were not influenced by their peer composition.

[Table 3.8 about here]

3.6 Conclusion & Discussion

The results of this study add to prior evidence on the effects of corequisite remediation reforms in Tennessee (Ran & Lin, 2022), Texas (Meiselman & Schudde, 2022; Miller et al., 2022), and City University of New York (Logue et al., 2019). While most of the existing evidence focuses on higher-scoring remedial students, we expand the evidence to include students with lower placement test scores. We found that although the magnitudes of the effects vary, placement into corequisite remediation led to substantially higher gateway course completion rates—a 76% improvement from prerequisite cohorts for math and a 40% improvement for English—for students across the placement test score distributions.

These positive effects on college-level credit accumulation were primarily the result of structural changes under the corequisite system. Corequisite remediation reduced the delay caused by lengthy prerequisite sequences, and students with lower test scores benefited the most from the mainstreaming effects. In addition, the structural reform also had diversionary effects: it replaced enrollment in remedial sequences with college-level courses. After the corequisite reform, remedial students accumulated similar amounts of college-level credits while enrolling in fewer courses overall, making the allocation of time and financial aid resources more efficient.

However, these positive effects on early college-level credit accumulation did not lead to improved downstream outcomes. We found that remedial students, particularly those with lower placement scores, were more likely to drop out and were less likely to earn short-term certificates. With more than 20 states and higher education systems in the nation adopting corequisite remediation, colleges need to continue identifying effective curriculum designs and pedagogical practices to help students with remedial needs succeed in college-level courses. The fact that college-level course pass rates and average grades became lower for corequisite cohorts than for prerequisite cohorts indicates that the concurrent academic learning support was

still not enough for some remedial students to succeed in college-level courses. In addition, with both on- level and remedial students, college-level classrooms have become more heterogeneous learning environments. These changes present new challenges for faculty to accommodate students with a wider range of academic needs.

A broader question for the field is whether remediation reform alone is expected to improve college completion. A series of recent studies found that remedial reforms focusing on structural changes had little impact on enrollment persistence or completion outcomes (Meiselman & Schudde, 2022; Miller et al., 2022; Ran & Lin, 2022). Together with similar results from the initiative to move remediation to high school (Kane et al., 2021), the literature in this area provided emerging evidence that the problems with traditional remediation models were not the primary drivers of low college completion rates. Interventions that showed positive effects on attainment outcomes incorporated components of curriculum reforms (Logue et al., 2019) or a holistic set of academic and financial support (Scrivener et al., 2015). Students may need ongoing academic and nonacademic support to translate the momentum gained from additional college-level coursetaking into improved persistence and completion outcomes.

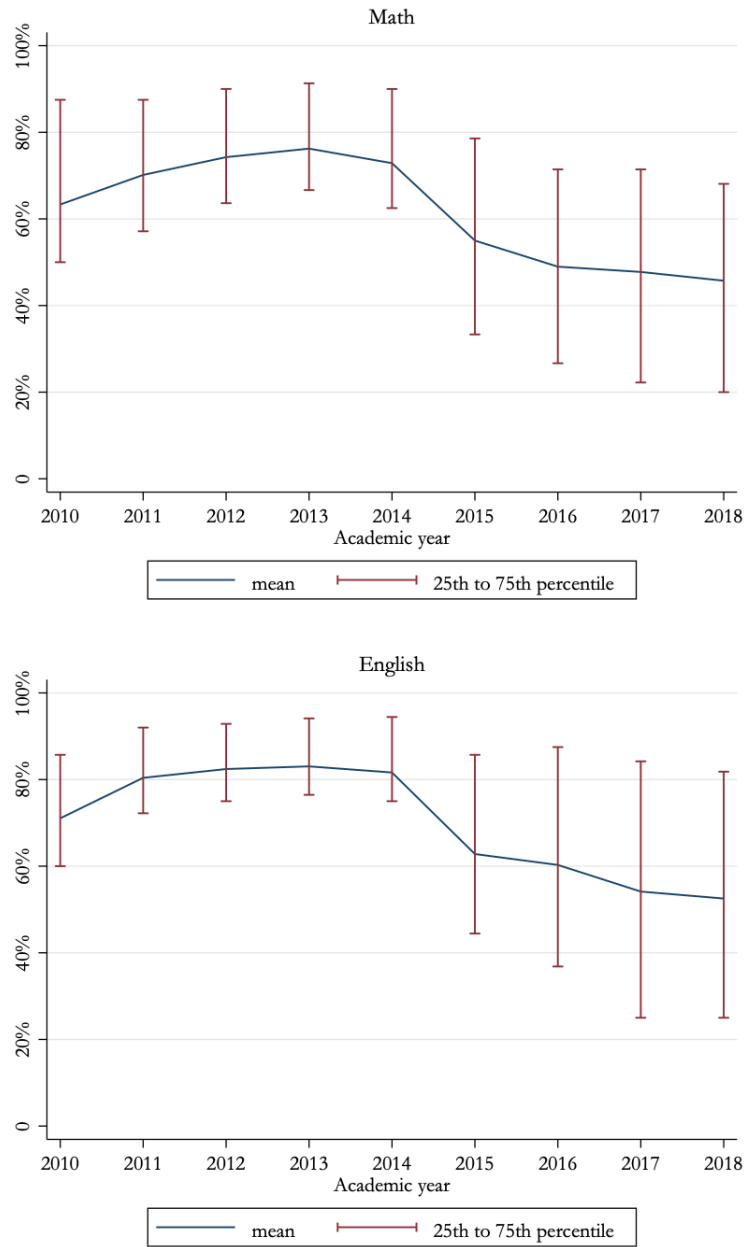


Figure 3.1: Proportions of on-level peers in gateway course sections by academic year

Note: These graphs show the average proportions of on-level peers in gateway course sections along with the 25 to 75 percentile ranges, by academic year. We define on-level peers as the students, other than oneself, who were placed into a gateway course section directly or had completed required remedial courses before enrolling in a gateway course section.

Table 3.1: Summary Statistics (2010–11 to 2018–19 Cohorts)

	Total	Prerequisite cohort	Corequisite cohort
Panel A. Background variables			
Female	0.543	0.551	0.536
Age at college entry	18.486	18.546	18.432
Race			
White	0.705	0.724	0.687
Black	0.195	0.198	0.192
Hispanic	0.053	0.039	0.066
Other race	0.047	0.039	0.054
International students	0.002	0.003	0.002
High school GPA	2.769	2.873	2.677
ACT score - Math	17.882	17.793	17.960
ACT score - English	18.645	18.729	18.570
ACT score - Reading	19.444	19.398	19.485
Referred to remediation	0.754	0.755	0.753
Panel B. Outcome variables			
<i>Gateway courses outcomes by Y1</i>			
Completed gateway Math	0.343	0.269	0.408
Completed gateway English	0.608	0.563	0.648
<i>Second college-level courses in math and English by Y2</i>			
Enrolled in a second math course	0.161	0.145	0.174
Completed a second math course	0.134	0.124	0.143
Enrolled in a second English course	0.465	0.438	0.488
Completed a second English course	0.382	0.369	0.394
<i>College-level credit accumulation by Y3*</i>			
# of credits enrolled	39.165	38.868	39.505
# of credits earned	29.647	29.700	29.586
<i>Persistence, transfer, and credential completion by Y3*</i>			
Dropout	0.575	0.530	0.626
Transfer to TN 4-year college	0.096	0.097	0.095
Earned any credential	0.129	0.104	0.158
N	91,511	42,904	48,607

Note: Descriptive statistics were based on students entering TBR community colleges during fall semesters in the year of high school graduation.

*Calculations for three-year outcomes were based on the 2010–11 to 2017–18 cohorts. The numbers of observations are 80,446 for the full analytic sample and 37,542 for the corequisite sample.

Table 3.2: Effects of corequisite remediation on first-year gateway course completion

	(1)	(2)	(3)	(4)
Panel A. DID estimates	Two-way fixed-effects estimates		Heterogeneity-robust estimates	
	Math	English	Math	English
Below cutoff * coreq	0.204*** (0.019)	0.228*** (0.030)	0.183*** (0.033)	0.236*** (0.040)
Baseline mean (SD)	0.269 (0.443)	0.563 (0.496)	0.269 (0.443)	0.563 (0.496)
N	91,511	91,511	341	341
Panel B. Event study estimates	Two-way fixed-effects estimates		Heterogeneity-robust estimates	
	Math	English	Math	English
Pre4 * below cutoff	0.002 (0.014)	0.003 (0.018)	-0.032 (0.029)	-0.032* (0.014)
Pre3 * below cutoff	0.003 (0.016)	0.023 (0.019)	-0.019 (0.025)	-0.010 (0.018)
Pre2 * below cutoff	0.002 (0.019)	0.025 (0.033)	-0.010 (0.030)	-0.022 (0.017)
Pre1 * below cutoff	0.098*** (0.015)	0.066 (0.029)	0.0826** (0.027)	0.008 (0.011)
Pre0 * below cutoff	0.240*** (0.020)	0.239*** (0.016)	0.222*** (0.029)	0.201*** (0.014)
Post1 * below cutoff	0.236*** (0.022)	0.247*** (0.024)	0.224*** (0.032)	0.214*** (0.027)
Post2 * below cutoff	0.211*** (0.013)	0.252*** (0.023)	0.218*** (0.026)	0.256*** (0.029)
Post3 * below cutoff	0.216*** (0.015)	0.269*** (0.014)	0.238*** (0.030)	0.280*** (0.028)
N	91,511	91,511	91,511	91,511

Note: The TWFE models controlled for students' demographic and pre-college academic characteristics in Table 1 Panel A, college fixed-effects, cohort fixed-effects, and high school fixed-effects. The heterogeneity-robust DID model was conducted at college-cohort-remedial status level since Stata packages `csdid` requires perfectly balanced panel. Standard errors are clustered at college-, cohort-, and high school-level. Robust standard errors are shown in parentheses. Pre-periods stand for the years before the corequisite implementation, and post-periods stand for the years after the implementation. Pre0 represents the year of the corequisite reform. We used pre5 as the reference group for all event-study specifications. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3.3: Effects of corequisite remediation on first-year gateway course completion by placement test scores

	(1)	(2)	(3)	(4)
	Math		English	
	Complete	Baseline Mean	Complete	Baseline Mean
Group 1	0.150*** (0.023)	0.025 (0.156)	0.326*** (0.049)	0.178 (0.382)
Group2	0.198*** (0.021)	0.092 (0.289)	0.232*** (0.024)	0.400 (0.490)
Group3	0.209*** (0.020)	0.238 (0.426)	0.136*** (0.023)	0.543 (0.498)
N	91,511		91,511	

Note: Results in this table are effects of placing into corequisite remediation for students by groups of placement test scores. All models controlled for students' demographic and pre-college academic characteristics shown in Table 1 Panel A, college fixed-effects, cohort fixed-effects, and high school fixed-effects. Standard errors are clustered at college-, cohort-, and high school-level. Robust standard errors are shown in parentheses. Group 1 (lowest scoring DE group) is defined as students whose ACT scores were 6 or more points below the cutoff. Group 2 represents students whose ACT scores were 3 to 5 points below the cutoff. Group 3 (highest scoring DE group) includes students whose scores were 1 or 2 points less than the college-ready criteria. Group 4, the reference group, includes students whose scores were at or above the cutoff. We used students' minimum test scores of writing and reading to define the test score subgroup for English.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3.4: Effects of corequisite remediation on subsequent college-level course outcomes (by end of Y2)

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
	Math			English		
	Enroll	Complete	Grades	Enroll	Complete	Grades
Below cutoff * coreq	0.056** (0.015)	0.040* (0.012)	-0.174** (0.047)	0.123*** (0.012)	0.076*** (0.011)	-0.193*** (0.035)
Baseline mean	0.145	0.124	2.412	0.438	0.368	2.431
(Baseline SD)	(0.352)	(0.329)	(1.372)	(0.496)	(0.482)	(1.336)
N	91,511	91,511	13,234	91,511	91,511	40,180
Panel B						
	Subgroup Analysis			Subgroup Analysis		
	Enroll	Complete	Grades	Enroll	Complete	Grades
Group 1* coreq	0.040 (0.020)	0.029 (0.016)	-0.424 (0.269)	0.151*** (0.020)	0.093** (0.019)	-0.282** (0.069)
Group2 * coreq	0.050** (0.015)	0.035* (0.011)	-0.179* (0.076)	0.123*** (0.011)	0.076*** (0.009)	-0.165** (0.042)
Group 3* coreq	0.064** (0.017)	0.046* (0.014)	-0.165** (0.042)	0.098*** (0.012)	0.059*** (0.011)	-0.188*** (0.035)
Baseline mean & S D						
Group 1	0.022 (0.147)	0.018 (0.134)	2.261 (1.453)	0.162 (0.369)	0.127 (0.333)	2.083 (1.312)
Group 2	0.058 (0.234)	0.048 (0.214)	2.154 (1.321)	0.304 (0.460)	0.244 (0.429)	2.137 (1.306)
Group 3	0.124 (0.329)	0.106 (0.308)	2.383 (1.324)	0.395 (0.489)	0.329 (0.470)	2.322 (1.305)
N	91,511	91,511	13,234	91,511	91,511	40,180

Note: All models controlled for students' demographic and pre-college academic characteristics shown in Table 1. Panel A, college fixed-effects, cohort fixed-effects, and high school fixed-effects. Standard errors are clustered at college-, cohort-, and high school-level. Robust standard errors are shown in parentheses. Group 1 (lowest scoring DE group) is defined as students whose ACT scores were 6 or more points below the cutoff. Group 2 represents students whose ACT scores were 3 to 5 points below the cutoff. Group 3 (highest scoring DE group) includes students whose scores were 1 or 2 points less than the college-ready criteria. Group 4, the reference group, includes students whose scores were at or above the cutoff. We used students' minimum test scores of writing and reading to define the test score subgroup for English. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3.5: Effects of corequisite remediation on credit accumulation, persistence, transfer and completion (by end of Y3)

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	Total number of credits		Enrolled	Earned	College-level credits		Enrolled	Earned	Left TN public system	Enrollment	Transfer to 4yr inst	Credential		
Panel A: All Students														
Below cutoff * coreq	-1.417**			-1.610**	2.512*	0.712			0.043**	-0.028				
	(0.356)			(0.399)	(0.751)	(0.485)			(0.010)	(0.012)				-0.030*
Baseline mean & SD	45.292			33.554	38.868	29.700			0.530	0.097				(0.009)
	(22.747)			(25.028)	(24.056)	(24.706)			(0.499)	(0.296)				0.104
N	80,446			80,446	80,446	80,446			80,446	80,446				(0.305)
														80,446
Panel B: Subgroup Analysis														
Group 1 * coreq	-1.131			-1.383	5.055**	1.865			0.039*	-0.044*				-0.061**
	(0.706)			(0.665)	(1.219)	(0.800)			(0.011)	(0.017)				(0.015)
Group 2 * coreq	-1.329**			-1.433**	2.735**	0.952*			0.047**	-0.030				-0.031*
	(0.352)			(0.340)	(0.598)	(0.346)			(0.011)	(0.013)				(0.009)
Group 3 * coreq	-1.822**			-2.144**	0.304	-0.618			0.041**	-0.016				-0.011
	(0.462)			(0.420)	(0.549)	(0.440)			(0.008)	(0.009)				(0.006)
Baseline mean & SD:														
Group 1	37.580			22.893	23.945	15.844			0.656	0.043				0.024
	(21.722)			(22.116)	(21.226)	(19.278)			(0.475)	(0.204)				(0.152)
Group 2	43.570			30.827	34.707	25.387			0.556	0.075				0.060
	(22.667)			(24.256)	(22.995)	(22.851)			(0.497)	(0.263)				(0.238)
Group 3	47.950			37.145	43.439	33.940			0.491	0.108				0.119
	(22.523)			(25.137)	(22.726)	(24.502)			(0.500)	(0.310)				(0.324)
N	80,446			80,446	80,446	80,446			80,446	80,446				80,446

Note: Results are based on analyses for students from 2010-11 to 2017-18 cohorts. All outcomes are tracked up to three years since initial term enrolled at TBR. All models controlled for students' demographic and pre-college academic characteristics shown in Table 1 Panel A, college fixed-effects, cohort fixed-effects, and high school fixed-effects. Standard errors are clustered at college-, cohort-, and high school-level. Robust standard errors are shown in parentheses. Group 1 (lowest scoring DE group): ACT 6+ points below cutoff. Group 2: 3-5 points below cutoff. Group 3: 1-2 points below. Group 4 (ref): at/above cutoff. English groups are based on minimum writing and reading scores. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3.6: Timing of first gateway course enrollment before and after corequisite reform by test score group

	Math		English	
	Prerequisite	Corequisite	Prerequisite	Corequisite
Panel A: Group 1 (6 points below the college-readiness threshold)				
Term 1	< 3%	24.12%	5.26%	75.62%
Term 2	4.98%	26.38%	23.22%	10.93%
Term 3	6.45%	3.79%	12.54%	< 3%
Term 4	6.45%	< 3%	7.56%	< 3%
Did not enroll by term 4	81.29%	43.00%	51.42%	11.24%
N	1,085	1,107	5,556	7,066
Panel B: Group 2 (3-5 points below the college-readiness threshold)				
Term 1	4.01%	33.06%	12.96%	79.37%
Term 2	11.27%	25.45%	43.35%	11.81%
Term 3	12.70%	4.95%	9.52%	< 3%
Term 4	10.03%	< 3%	4.07%	< 3%
Did not enroll by term 4	62.00%	32.06%	29.55%	7.28%
N	17,438	18,655	9,682	11,141
Panel C: Group 3 (1-2 points below the college-readiness threshold)				
Term 1	8.57%	45.09%	26.58%	80.19%
Term 2	25.50%	23.47%	45.64%	12.00%
Term 3	10.32%	5.35%	6.33%	< 3%
Term 4	9.5%	3.04%	< 3%	< 3%
Did not enroll by term 4	42.38%	23.05%	19.09%	6.59%
N	11,198	13,361	6,543	7,718

Note: Group 1 (lowest scoring DE group) is defined as students whose ACT scores were 6 or more points below the cutoff. Group 2 represents students whose ACT scores were 3 to 5 points below the cutoff. Group 3 (highest scoring DE group) includes students whose scores were 1 or 2 points less than the college-ready criteria. Students whose placement test scores were above college-readiness threshold are not included in this table.

Table 3.7: Number of developmental credits enrolled and earned before and after corequisite reform by test score group

	Panel A: Prerequisite students					
	Enrolled remedial credits			Earned remedial credits		
	By end of Y1	By end of Y2	By end of Y3	By end of Y1	By end of Y2	By end of Y3
Group 1	11.768	13.348	13.634	6.218	6.918	7.048
Group 2	7.880	8.699	8.853	4.968	5.362	5.437
Group 3	4.109	4.441	4.516	2.970	3.156	3.203
	Panel B: Corequisite students					
	Enrolled remedial credits			Earned remedial credits		
	By end of Y1	By end of Y2	By end of Y3	By end of Y1	By end of Y2	By end of Y3
Group 1	6.603	7.163	7.230	3.374	3.628	3.660
Group 2	4.553	4.846	4.885	2.913	3.078	3.102
Group 3	2.387	2.521	2.541	1.705	1.785	1.797

Note: Calculations were based on students whose placement test scores were below college-readiness threshold. We imputed zero developmental credits enrolled for students who were placed into but did not enroll in any remedial courses and students who left TBR community colleges. Group 1 (lowest scoring DE group) is defined as students whose ACT scores were 6 or more points below the cutoff. Group 2 represents students whose ACT scores were 3 to 5 points below the cutoff. Group 3 (highest scoring DE group) includes students whose scores were 1 or 2 points less than the college-ready criteria.

Table 3.8: Association between proportion of on-level peers and first-year gateway course completion rates

	(1)	(2)	(3)	(4)	(5)	(6)
	All gateway enrollees	Math enrollees above college-level cutoff	Enrollees below college-level cutoff	All gateway enrollees	English enrollees above college-level cutoff	Enrollees below college-level cutoff
Post corequisite cohorts	0.247*** (0.042)	0.076 (0.098)	0.241*** (0.035)	0.179* (0.091)	0.053 (0.080)	0.171* (0.084)
% on-level students	0.302*** (0.104)	0.337*** (0.146)	0.221*** (0.091)	0.221* (0.118)	0.160 (0.088)	0.273*** (0.113)
Post % on-level	-0.351*** (0.099)	-0.201 (0.149)	-0.267*** (0.100)	-0.268* (0.128)	-0.038 (0.094)	-0.330*** (0.120)
N	49,161	21,579	27,582	77,590	41,856	35,734

Note: Each column presents results from a regression using a difference-in-differences set up, where the main predictors include an indicator for post-corequisite cohorts, proportion of on-level students in the gateway course section, and the interaction term between them. The models also controlled for students' demographic and pre-college academic characteristics shown in Table 1 Panel A, college fixed-effects, cohort fixed-effects, and high school fixed-effects. Standard errors are clustered at college-, cohort-, and high school-level. Robust standard errors are shown in parentheses. Columns 1 and 4 are results for all students who enrolled in gateway courses during the first year, columns 2 and 5 are results for students whose placement test scores were above college-level cutoff, and columns 3 and 6 are results for students who tested below college-level.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Chapter 4

PAPER THREE: THE ARCHITECTURE OF EXPECTED WAGE GAPS: BETWEEN- AND WITHIN-SCHOOL SOURCES OF CAREER EDUCATION INEQUALITY (WITH DR. KENNETH A. SHORES AND ARIELLE LENTZ)

Abstract School-level variation in early career preferences likely shapes wage inequality. Using Delaware administrative data and Bureau of Labor Statistics occupational wage data, we characterize expected wage inequality in Career and Technical Education by analyzing how student demographics influence selection into programs of study with different expected wages. Multilevel mixed-effects modeling reveals substantial wage gaps across student subgroups, with traditionally disadvantaged students concentrating in lower-wage pathways. Decomposition analyses show gender expected wage gaps stem from within-school factors, while racial and socioeconomic gaps arise from between-school differences. Student segregation into programs contributes more to these inequalities than schools' program offerings. The findings suggest targeted interventions: addressing within-school practices for gender gaps and between-school resources for racial and socioeconomic disparities.

4.1 Introduction

In the United States, an adult’s earned wages directly impact their wellbeing, health, and stability (Chetty et al., 2016; Killingsworth, 2021). However, wages vary greatly between fields of employment, and even within employment fields, wage disparities persist across genders, races, abilities, and language status (Blau & Kahn, 2017; McCall, 2001; O’Neill & O’Neill, 2006). An ongoing priority for the country is to embed greater equity in the workplace, as workers of color, workers with disabilities, workers who have immigrated, and workers who identify as non-male genders face wage inequities requiring redress (U.S. Department of Labor, 2024).

Much research on wage inequities focuses on actual labor market outcomes, such as whether certain demographic groups earn differential wages or work in particular fields (Bobbitt-Zeher, 2007). Another substantial body of research highlights relationships between college-going behaviors and labor market outcomes, and how these differ between demographic groups (Renzulli et al., 2006; Sloane et al., 2021). The emphasis on college and labor market outcomes is unsurprising, given the close timing of these events. However, we recognize that these different outcomes likely start much earlier, before the disparities in the labor market become apparent.

Indeed, at least by high school, an individual begins to select coursework and make plans in anticipation of a future career that potentially influences the wages they earn as an adult (Altonji et al., 2012; Card & Payne, 2021; Nagy et al., 2008; Sadler et al., 2012). Many high schools have diversified their course offerings to help students prepare for careers, such as introducing advanced placement (AP), international baccalaureate (IB), and career and technical education (CTE) courses. However, a high schooler’s course enrollment to support their career aspirations is influenced by external factors, such as the courses available to students and the adults guiding them (Altonji et al., 2012; Aschbacher et al., 2010; Dick & Rallis, 1991; Wahl & Blackhurst, 2000), as well as internal factors such as their family backgrounds,

beliefs about careers and gender norms, sense of self-efficacy, knowledge and skills (Cho-Baker et al., 2021; Correll, 2001; Evans & Diekmann, 2009; Mau & Li, 2018).

We therefore consider it worthwhile to understand how heterogeneous and inequitable preferences for careers form in high school, independently of actual wages earned. This question bears importance partly because it allows for the separation of wage inequities arising from the direct effects of labor market participation (e.g., as arising from discrimination or sector- and firm-specific preferences) and those arising from exposure to careers and training that takes place in high school. As such, findings from this inquiry could speak to interventions that can apply earlier in an individual's development. Lastly, career-based course offerings (i.e., CTE programs) create a new type of social stratification within schools, distinct from the forms of tracking that have been traditionally studied (Gamoran, 1992). Indeed, as CTE has expanded to encompass multiple professional sectors, there is increasing need to document stratification occurring within CTE programs and the factors associated with this stratification (Giani, 2019).

The extent to which variability and inequity in wage selection begins to form in high school is therefore an important question, but one not easily answered due to data limitations. Typically, data on this question have been limited to surveys that ask about students' career aspirations. Holding aside the difficulty of collecting comprehensive and annual survey data on aspirations (e.g., as is collected via the High School Longitudinal Study), aspirations are not necessarily indicative of behavior. In contrast, a CTE career pathway requires successfully completing multiple courses in a specific field of study to become a CTE concentrator. And, because CTE course completion determines a concentration in a career pathway, recurring annual descriptions of this behavior are readily available with student-level administrative data.

CTE courses and students' career pathway concentration therefore provide a different type of social stratification that manifests relatively early in a student's

professional career, are (potentially) substantially unequally distributed, suitable for early policy intervention, and consequential for students' long-run earnings potential, as CTE concentration has been shown in multiple studies to causally increase labor force participation and earnings (Ecton & Dougherty, 2021; Stevens et al., 2019). Our goal in this study is therefore to quantitatively describe the degree of social stratification within CTE programs of study, focusing on group-level inequality between and within schools.

To do this, we operationalize a concept we call “expected wages,” which are based on the wages assigned to occupational codes from the Bureau of Labor Statistics, which can then be attached to CTE programs of study. In effect, for each CTE program of study, we can know its expected wage, and these expected wages then form the dependent variable for our analysis.

1. What are the expected wage gaps among student subgroups based on their selected career cluster?
2. How much of these differences in expected wage gaps are due to between versus within school differences?
3. What school-level factors predict variation in expected wage gaps?

Our study context is Delaware, which is a fruitful place to conduct this analysis, as about 90% of students participate in CTE and 60% of Delaware high school graduates are CTE concentrators, as compared to the 49% of students participating in CTE nationally (NCES2019 ; Association for Career and Technical Education, 2021; of Education, 2023). Using course data, we identify the CTE program of study that each of these CTE concentrators in the state belong to and link their individual program to aggregate wage data as described above. Our focus is on wage gaps—understood in this context to mean expected wage gaps based on program choice—between specific groups of students (focusing on differences between

gender (coded male or female as recorded by the Delaware Department of Education) race/ethnicity, and economic disadvantage) and the observed factors that are associated with these wage gaps.

4.2 Research Context: Career Technical Education in Delaware

Since 1999, the US Department of Education has worked with the Office of Vocational and Adult Education to create a uniform framework, called the National Career Clusters Framework, to guide CTE implementation across the country (Advance CTE, 2023). This Framework puts forth 16 “career clusters” that define common knowledge and skills aligned to broad occupational groupings (e.g., Architecture & Construction, Health Science, Hospitality & Tourism) (Advance CTE, 2023; Delaware Department of Education, 2017). Within the career clusters are “career pathways,” or distinct sets of courses that prepare students for a particular career (e.g., Design & Pre-Construction, Health Informatics, Restaurant & Food Services) (Advance CTE, 2023; Delaware Department of Education, 2017). States are also encouraged to offer “programs of study” (POS) which are specific sequences of courses designed to prepare students for field-specific higher education or the workforce (Delaware Department of Education, 2017). Students are then able to earn the distinction of a “concentrator” upon graduation, meaning they took enough coursework to specialize in a particular POS.

In 2018, the Strengthening Career and Technical Education for the 21st Century Act (Perkins V) enabled states to adjust CTE programs to reflect local needs, align with local job markets, and establish definitions for “concentrating” that accurately document processes and outcomes aligned to their CTE programs (ExcelinEd, 2018). Perkins V provided a baseline definition to identify students as concentrators, including any secondary school level student who has completed at least 2 courses in a single CTE POS (Carl D. Perkins Career and Technical Education Act of 2006, 2006). While this definition serves as a basis for state guidance, states are then able

to modify this definition to align with their data collection efforts. In Delaware, a student is considered a concentrator if they participate in two or more sequenced CTE courses within a POS, and Local Education Agencies (LEAs) are responsible for making this distinction for students (Delaware Department of Education, 2020).

4.3 Conceptual Framework: Expected Wages as a Measure of Educational Stratification

In the study of social stratification within educational processes, researchers have traditionally relied on three primary measures: achievement scores (Reardon et al., 2019, 2014), career aspiration surveys (Signer & Saldana, 2001), and categorical inequalities (Domina et al., 2017; Shores et al., 2020). While each of these measures provides valuable insights, they also have limitations in capturing the full picture of how educational experiences translate to future economic outcomes. To address these limitations and offer a novel perspective on educational stratification, we propose a fourth measure: "expected wages."

The traditional measures, while informative, each present specific challenges. Achievement scores, though indicative of academic performance, are less directly linked to career preferences and potential earnings. Career aspiration surveys reflect student preferences but may not align with actual behaviors or realistic outcomes. Categorical inequalities capture important distinctions in educational experiences, such as enrollment in advanced courses, but often rely on discrete categorizations rather than continuous measures.

Our proposed measure of expected wages combines elements of preference, behavior, and potential economic outcomes, offering several advantages over traditional approaches. First, by using Career and Technical Education (CTE) program enrollment data, it reflects actual student choices rather than just aspirations, providing a behavioral basis for our analysis. Second, it offers a projection of potential

earnings, bridging the gap between current educational experiences and future economic outcomes. This future-oriented approach allows us to examine how early career preferences and educational choices may contribute to wage inequalities.

Unlike categorical inequalities, expected wages offer a continuous scale of potential stratification. This continuous measure provides a more granular view of disparities, allowing for a more detailed analysis of the factors contributing to educational and economic inequalities. Furthermore, the expected wages measure has significant policy relevance, allowing educators and policymakers to assess and address inequalities in real-time, without waiting for long-term labor market outcomes.

The concept of expected wages aligns with the broader understanding of schools as institutions that prepare students for future economic roles. It enables us to examine how early career preferences and educational choices may contribute to wage inequalities, moving beyond simple academic achievement metrics. Additionally, this approach allows us to identify and address stratification within CTE programs, shifting the focus from the traditional CTE versus non-CTE dichotomy to a more nuanced understanding of within-program disparities.

By focusing on expected wages, we can also better understand the mechanisms through which educational experiences translate into economic outcomes, offering a more comprehensive view of social stratification in education. This approach provides a more immediate feedback loop for policymakers seeking to reduce inequality in high school settings. It allows for the identification of potential wage gaps early in a student's educational journey, potentially informing interventions or policy changes to address these disparities before they manifest in the labor market. Our specific focus on school-level factors associated with expected wage inequality bolsters this idea.

In summary, the introduction of expected wages as a measure of educational stratification offers a novel and potentially powerful tool for researchers and policymakers. By combining elements of student choice, projected economic outcomes,

and a continuous scale of measurement, this approach addresses many of the limitations of traditional measures. As we apply this framework to our analysis of CTE programs in Delaware, we anticipate gaining new insights into the formation of wage inequalities and the role of educational institutions in shaping future economic outcomes.

4.4 Data

We use two datasets to address our research questions: students' administrative records and aggregate occupational wages from the Bureau of Labor Statistics (BLS). First, the Delaware Department of Education provided the student administrative dataset that we use to construct our sample and act as independent variables. This longitudinal dataset included 57,766 high school students, graduating from 2017 to 2021 alongside information on student demographics (i.e., gender, race/ethnicity) as well as their participation in special programs, such as Free and reduced-price lunch (FRPL), Individualized Education Programs (IEP) and English Language Learner (ELL) programs. The data also includes CTE course enrollment, academic performance, the high school attended, and concentrator status as identified by LEA. We define concentrator status using course records, determining that a student is a concentrator in CTE POS j if the student completed two courses in a given POS with at least one course at or above a level 2 (see Huang et al. (2024) for additional details). We then implement several sample restrictions to assess whether test scores play a key role in determining students' benefits in CTE POS participation. We retain students with 8th-grade English Language Arts (ELA), math, and science scores who graduated from a Delaware public high school and meet the requirements of the Perkins V definition of concentrators. This restriction captures 33,275 students, or about 57% of the original sample.

Second, we use occupational employment and wage statistics from the BLS to generate our outcome variable which is the mean expected wage of students selected

POS. We link BLS data to DDOE student administrative data using CIP SOC Crosswalk provided by the National Center for Education Statistics . CIP SOC crosswalk is a dataset that matches the 6-digit 2020 Classification of Instructional Program (CIP) and 2018 Standard Occupational Classification (SOC). If multiple occupational codes are assigned to a POS, we calculate the mean wage of the assigned occupations. In our data, we are able to link 101 POS codes to SOCs out of 185 POSs provided by schools in Delaware, giving us a final analytic sample of 18,329 students, comprising about 31% of the original sample size.

Given the country’s efforts to improve workplace equity for workers of color, workers with disabilities, non-native English speakers, and workers who identify as non-male genders, we descriptively show several subgroups of interest in Table 4.1, including male and female students, and white, Black, Hispanic, and Asian students. We additionally provide information comparing students participating in special programs (FRPL, IEP, and ELL) with those who do not participate in each respective program. We show proportions of student subgroups participating in each career cluster, alongside that career cluster’s expected wage. While the statewide sample consists of 51% male and 49% female students, the share of male and female students varies significantly by career cluster. The majority of our students are white (50%), followed by black (18%), Hispanic (16%), and Asian (3%).

An advantage of the descriptive statistics shown in Table 4.1 is that it is easy to see the sources of expected wage inequality, which stem from two mechanisms: inequality in expected wages by cluster and differential participation in cluster by student group. For example, Cluster 9 has the greatest representation of black students (48%), yet the lowest mean expected wages across all clusters (\$32,197). Cluster 17 has the largest representation of Hispanic students (38%) and a moderate expected wage of \$79,823. However, Cluster 11 has the greater representation of white and Asian students (62% and 10% respectively), and the highest expected wages across all clusters (\$106,194). From this aggregated data flows substantial

inequality in expected wages between White and Asian students on the one hand, and Black and Hispanic students on the other. A similar set of results is observable for gender and other groups.

These aggregate data mask variation in expected wages within clusters (i.e., at the POS level) and between schools. For example, in Cluster 1, the expected wage varies across POS from \$26,505 to \$156,110. These descriptive results, however, lay the groundwork for the quantitative analyses described below.

[Table 4.1 About Here]

To further illustrate the mechanisms undergirding expected wage inequality, we focus on gender and, in Figure 1, plot the expected wage assigned to a POS against the proportion of females concentrating in that POS for individual schools in the three public districts in the state. The slopes of these lines indicate increases in the expected wage inequality, with positive slopes meaning that more females are represented in high wage POSs. One can see (Figure 4.1 Panel A) that these slopes are variable between schools within districts and between districts, meaning that a formal analysis decomposing between and within district variance could prove fruitful.

[Figure 4.1 About Here]

4.5 Methods

To answer RQ1, which aims to estimate the expected wage gap between student subgroups, we employed a multilevel mixed-effects model that accommodates both fixed and random effects. Our analysis begins with an examination of the relationship between students' CTE expected wages and their demographic backgrounds. The multilevel mixed-effects model can be written as follows:

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_{gj}D_{ij} + X_{ij} + \delta_c + \gamma_{ij} \quad (4.1)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j} \quad (4.2)$$

$$\text{Level 2: } \beta_{gj} = \gamma_{g0} + u_{gj} \quad (4.3)$$

Here, Y_{ij} denotes the outcomes of interests, mean of expected wage from the POS-connected occupations. D_{ij} is a binary variable showing the main explanatory variable of our interests, which includes students' gender, race/ethnicity, FRPL, ELL, and IEP status. Because we are interested in inequalities between specific groups of students – for example, between Black and White students – we estimate these models sequentially at the subgroup-type level. Specifically, we estimate models for (i) gender, (ii) race/ethnicity, (iii) economic disadvantage, (iv) ELL status, and (v) IEP status. Estimating at the subgroup-type level allows us to provide the true difference in expected wages and not the conditional differences (e.g., conditional on economic disadvantage). X_{ij} controls for student's 8th grade test scores, which are included in separate models, δ_c represents cohort-fixed effects. γ_{ij} is the error term.

One benefit of the multilevel mixed-effects model is that it allows for random intercepts and random coefficients for the explanatory variables of our choice to observe school-specific variations in the correlations we are interested in. These relationships are shown through β_{0j} and β_{gj} . Level 2 equations divide school-specific intercepts and slopes into across-school averages and school-specific parts. γ_{00} shows the average of the school means, and u_{0j} indicates the unique intercept for school j . Similarly, γ_{g0} is the average difference in expected wages (e.g., male vs female expected wage) for the entire sample while u_{gj} denotes the school-specific difference in expected wages between the focal group (e.g., male students) and the reference group (e.g., female students).

Next, we use Reardon's (2008) gap decomposition strategy to identify how much of the estimated wage gap is due to between versus within school differences. Reardon (2008) combines two decomposition strategies proposed by Fryer Jr and

Levitt (2004) (henceforth FL) and Hanushek and Rivkin (2006) (henceforth HR). Reardon (2008) shows that the Black and White gap (in his case achievement and in our case the expected wage) can be divided into three parts: the unambiguous between-school gap, the unambiguous within-school gap, and an ambiguous portion. To decompose the expected wage gap between the focus group (male, black, Hispanic, ELL, FRPL, and IEP students) and the reference group (female, white, non-ELL, non-FRPL, and non-IEP students), we use the model from Reardon (2008). First, we estimate the following regression to measure the focus-reference group expected wage gap:

$$Y_{ij} = \beta_0 + \beta_1 D_{ij} + \beta_2 p_j + \varepsilon_{ij} \quad (4.4)$$

where D_{ij} is a binary variable indicating student i 's demographics in school j , and p_j is the proportion of students in school j who falls into our focus group. From this model, we estimate the average focus-reference group expected wage gap as follows:

$$\text{Wage gap} = \hat{\beta}_1 + \hat{\beta}_2 (\underline{p}_j^f - \underline{p}_j^r) \quad (4.5)$$

We can rewrite the gap again as follows:

$$\text{Wage gap} = \hat{\beta}_1 \left[1 - (\underline{p}_j^f - \underline{p}_j^r) \right] + (\hat{\beta}_1 + \hat{\beta}_2) (\underline{p}_j^f - \underline{p}_j^r) \quad (4.6)$$

$\hat{\beta}_1 \left[1 - (\underline{p}_j^f - \underline{p}_j^r) \right]$ is the within school gap, and $\hat{\beta}_2 (\underline{p}_j^f - \underline{p}_j^r)$ is between school gap. $\hat{\beta}_1 (\underline{p}_j^f - \underline{p}_j^r)$ is the proportion of the gap which comes from the interaction of between and within school components, making it ambiguous.

Finally, to answer RQ3, which investigates the school-level factors driving school-specific variation in the expected wage gap, we first categorize school-level factors into "program availability characteristics" and "CTE segregation patterns."

These factors map directly onto our analyses of between — versus within-school inequality. Program availability characteristics — measured by the number and average expected wages of POSs offered — could explain between-school gaps if certain schools systematically offer fewer or lower — wage programs. Meanwhile, CTE segregation patterns speak to within-school inequality by examining how different student groups distribute across available programs even within the same school. For program availability characteristics, we consider the number of POSs offered in the school and the mean expected wage of all POSs in the school. These metrics show the diversity of POS offerings within each school and indicate the extent to which higher-wage programs are accessible to students.

For CTE segregation, we use the Gini-Simpson diversity index, which measures how students are distributed across POS options within each school. To illustrate how this index captures meaningful variation in student sorting, consider three schools from our sample, each offering 6 POSs in 2017. In the first school, 87% of female students concentrated in a single POS while male enrollment was more evenly distributed, ranging from 2% to 52% across programs. In the second school, we observe the opposite pattern: female enrollment ranged from 14% to 54% across programs, while 88% of male students concentrated in a single POS. The third school showed relatively even gender distribution, with both female and male enrollment ranging from 2% to approximately 50% across programs.

To analyze these patterns systematically, we calculate the Gini-Simpson diversity index for each demographic group (i.e., gender, race/ethnicity, FRPL status, ELL status, and IEP status) and determine the diversity gap between focus and reference groups. The index for group d in school s (λ_{ds}) is calculated as follows: $\lambda_{ds} = 1 - \sum_{k, \text{ if } i \in d \in s} p_k^2$, where p_k indicates the proportion of students in POS k . The index equals zero if all students in the school enroll in one POS, and the index equals $1/k$ when all students in the school are equally distributed into k POS. The diversity gap is calculated by subtracting the reference group's index

from the focus group’s index. A positive gap indicates the focus group is more evenly distributed across POSs than the reference group. In our example schools above, the gender diversity gaps are 0.40, -0.41, and 0.04, respectively, capturing the substantial variation in gender segregation patterns across schools.

4.6 Results

4.6.1 Research Question One: Magnitude and Distribution of Inequality

We start by estimating unconditional wage gaps, which describe average inequality in expected wages for the state. We implement these models pairwise to avoid controlling for subgroup characteristics that might attenuate true group-specific differences (e.g., estimating Black-White expected wage differences controlling for economic disadvantage). The adjusted pairwise difference in wages between male and female students (Table 4.2, Column 1, Panel A) is \$2,641.37 ($p < 0.001$; 95% CI \$2,157.99 - \$3,124.75), meaning that males are expected to earn \$2,641.37 more than females. In Panel E, Black students are expected to earn \$2,659.05 less than white students ($p < 0.001$, 95% CI \$2,089.39 - \$3,228.71), and Hispanic students are expected to earn \$4,948.63 less than white students ($p < 0.001$, 95% CI \$4,261.07 - \$5,636.19). For students receiving FRPL, IEPs, ELL services (Panels B—E, respectively), gap magnitudes range from -\$2,338 to -\$7,713.

In Column 2, we incorporate a random intercept in the model; doing so allows us to test whether the average wage varies among schools and provides estimates comparable to fixed effects models, yielding expected wage inequalities roughly interpretable to the average within school expected wage gap. Controlling for between school variation in average wages changes gap magnitudes. For gender, the gap narrows only slightly to \$2,122.91 ($p < 0.001$, 95% CI \$1,677.25 - \$2,568.57), but racial/ethnic expected wage gaps change more dramatically, as Black and Hispanic wage differences are no longer statistically different from zero and are less than 20

percent the size of the total gap. Controlling for between school differences similarly attenuates expected wage gaps for students receiving FRPL, IEPs, and ELL services.

In Columns 3 and 4, we run the same models as shown in Columns 1 and 2 but control for 8th grade test scores. The influence of test scores on expected wage gaps is mixed. Gender gaps remain roughly the same (column 3; \$3,515.78; $p < 0.001$, 95% CI \$3,020.81 - \$4,010.75), which narrows slightly when including a random intercept (Column 4; $p < 0.001$, 95% CI \$2,241.01 - \$3,162.27). Black-White expected wage gaps reverse directionality, meaning that Black students with similar test scores as White students select into career POS with expected wages that are \$1,244 greater than White students ($p < 0.05$, 95% CI \$652.39 - \$1,836.17); including a school-level random intercept has little effect beyond the test score. For Hispanic-White gaps, the gap shrinks by more than half to -\$2,032.78 ($p < 0.001$, 95% CI \$1,339.16 - \$2,726.40) and reduces to \$219 after including school-level random intercepts but is no longer statistically significant. Controlling for test scores similarly attenuates expected wage gaps for students receiving FRPL, IEPs, and ELL services, in some cases eliminating the expected wage gap entirely.

4.6.2 Research Question Two: Within or Between School Factors

The descriptive evidence from Figure 4.1 Panel A showing between school differences in the gender composition of different POS, coupled with the attenuating effect of school-level random intercepts on expected wage gaps shown in Table 4.2, reveals the importance of school-level influences on the generation of expected wage inequality. We investigate the influence of schools as a source of inequality in two ways. First, we modify Equation 2 by including school-specific random slopes, which provide school-specific estimates of the expected wage gap, with and without controlling for test scores. These estimates tell us how much within-school expected wage gaps vary across the state, which are useful because, if there is variation across

schools, it suggests school-level factors can contribute to or ameliorate expected wage inequality. Should these school-level expected wage gaps be similar to controls for 8th grade test scores – an important source of student-based selection into career paths – this would provide additional evidence about the importance of schools, which would then implicate school-based policy solutions.

Table 4.3 presents results from this analysis. Including random slopes greatly attenuates the average difference in expected wage gaps for Black and Hispanic students relative to White students (Table 4.3, Column 1, Panel E) and increases the imprecision of the gender wage gap, though does not affect the coefficient. Expected wage differences for students qualifying for FRPL, IEPs, or ELL services are less affected. However, our main interest is in the variance of these components, which we transform to standard deviation (SD) units. Here we see confirmation that school-level factors are an important contributor to expected wage inequality: there is substantial variation in the expected wage across schools. For example, the SD of the random slope for gender is \$12,509 (also visualized in Figure 1, Panel B), meaning that if expected wage gaps by gender are normally distributed among schools, about 68% of schools have expected wage gaps falling within -\$9,220 to \$15,798. Though gender has by far the greatest variability, other gaps, especially relative to the average, vary as well. For example, the SD of the random slope for Black-White expected wage gaps is \$1,118, meaning that for about 68% of schools the Black-White expected wage gap falls within -\$1,702 to \$534, and for Hispanic-White expected wage gaps that interval encompasses -\$4,087 to \$2,169.

The variance in expected wage gaps across schools remains virtually unchanged when controlling for 8th grade test scores. The school-specific random slopes (corresponding to the “shrunk” Empirical Bayes estimate from Equation 3) show correlations of at least 0.99 with and without test score controls for each subgroup comparison. This extremely high correlation indicates that the relative ranking of schools in terms of their wage gaps is preserved—if School A had a larger

wage gap than School B before controlling for test scores, it still does after controlling for them. This persistence in the ordering of schools suggests that wage gaps between demographic groups at the school level are not primarily driven by differences in student academic preparation, but rather by school-level factors that generate these inequalities.

In Table D.2, we evaluate the robustness of this correlation by estimating school-specific wage gaps conditional on 8th grade test scores using ordinary least squares regression, coarsened exact matching (CEM) (Blackwell et al., 2009) and entropy balancing (Hainmueller & Xu, 2013), separately. These school specific estimates are noisier, and we expect the correlation to the unconditional models to attenuate, but they remain large, ranging between 0.78 to 0.99. Thus, we conclude that school-level variance in the expected wage gaps is largely uncorrelated with 8th grade test scores.

[Table 4.3 About Here]

Second, we decompose the overall wage gaps into their between and within school components. This decomposition answers a different but related question: how much of the total wage gap between groups stems from differences in the schools they attend (i.e., the between component) versus unequal outcomes within the same schools (i.e., the within component). While the multilevel model shows how wage gaps vary across schools, the decomposition tells us how much this cross-school variation contributes to total inequality relative to within-school processes.

Results are presented in Table 4.4. For gender, 83% of the expected wage gap occurs within schools. This result corresponds to results discussed above, with school-level variation in the expected wage gap being exceptionally high between male and female students. In contrast, for Black-White and Hispanic-White expected wage gaps, 66% and 81% of the variation occurs between schools, respectively, meaning that most of the expected wage gap for racial/ethnic inequality is due to

differences among schools in average expected wage offerings and the segregation of Black and Hispanic students into those schools. This result was foreshadowed when we showed that including a school-specific random intercept, controlling for average differences in expected wages between schools, greatly attenuated the gap for Black and Hispanic students. For students receiving FPRL and ELL services, most of the gap occurs between schools, whereas for students with IEPs, the majority of the gap occurs within schools.

Controlling for test scores (Table 4.4, Panel B) has no effect for gender gaps, as 85% of the gap remains due to within school factors. For Black-White and Hispanic-White gaps, the between school components continue to contribute to inequality disfavoring Black and Hispanic students – between school segregation and school-level differences negatively affect those expected wage gaps conditional on test scores. However, for Black students, because the gap is now reversed, the within-school component (\$2,100) is larger than the total gap (\$1,612); this means that if only within-school factors were considered, the gap favoring Black students would be even larger. The negative between-school component (-\$776) is partially offsetting this within-school advantage. The pattern is largely the same for Hispanic-White expected wage gaps, though the average gap remains negative: on average, within schools, the gap favors Hispanic students but the between-school component is large and negative.

[Table 4.4 About Here]

There is an apparent tension in the results presented above, as we observe large between-school variation in school-specific gender gaps, but the between-school component of the gender-expected wage gap is trivial. How can this be? First, recall that the expected wage gap decomposition is based on between school segregation; differences in the assignment of student groups across schools is a necessary feature of between school inequality. For gender, there is little segregation, as most schools

are roughly equally composed of male and female students. Second, the random slopes model captures variation in gender wage gaps that arises from school-specific contextual factors, such as differences in school culture, resources, or teacher practices, which may affect how male and female students experience schooling differently within each school, even if the overall gender composition across schools is similar.

4.6.3 Research Question Three: Predictors of School-Level Gaps in Wage Inequality

We identify school-level differences as mechanisms for expected wage inequality. These differences come from two sources, either variation in within-school practices that cause variation in school-specific expected wage inequality or average differences across schools in expected wages that drive expected wage inequality because of segregation. Our final research question seeks to identify potential mechanisms for this school-level variation. We consider program availability characteristics and CTE segregation patterns as two potential factors.

Of the hypotheses considered, only CTE segregation is a consistent predictor of expected wage inequality. As mentioned above, CTE POS segregation is measured using the Gini-Simpson diversity index, which captures how concentrated a focal group (e.g., male students) is across POS in a school compared to a reference group (e.g., female students). In Table 4.5, we see that on average, for all subgroups, the diversity index gap positively predicts group-level expected wage differences. For example, in schools where females are more concentrated in singular POS compared to male students (i.e., when male diversity is greater than female diversity across POS), the male-female expected wage gap grows. Specifically, for students attending a school where the male diversity index is 1-unit higher in the diversity index than it is for male students, male students concentrated in POS with expected wages that are \$23,911 greater than female students. A one-unit change is about 5.56 SD of variation, so this corresponds to about \$4,304 in expected wage differences for a 1

SD change in the diversity gap. These magnitudes and statistical significance are strikingly similar across all subgroups.

Conversely, when the number of POS offered increases or when the average expected wage increases, there is no consistent pattern (signs change across subgroup comparison) and coefficients are often not statistically distinguishable from zero. For gender, when the average expected wage gap increases by \$1.00, the gender wage gap increases by \$0.20; no other gaps are affected by this variable. For the number of pathways offered, expected wage gaps for gender and race/ethnicity do not vary. FRPL and IEP expected wage gaps narrow with additional pathways, whereas ELL expected wage gaps increase.

We run an alternative model that interacts the number of POS offered in the schools with the diversity gap within schools. The results (Table D.3) show that the increase in the number of POS options raises the expected earnings of Black and low-income students when these students are more widely spread across POSs offered in school. In Table D.4, we include the mean expected wage of all POSs in the school, and the results suggest that its association with expected wage gap is marginal. This again confirms that the schools' program availability characteristics is not consistently associated with the expected wage gap between student subgroups and that what seems to drive cross-school variation in expected wage inequality is the segregation of students in CTE programs of study.

[Table 4.5 About Here]

4.7 Policy Applications

These results lend themselves to potential policy actions. Importantly, our findings allow us to tailor interventions for group-specific inequalities distinguishing between within-school inequality and those inequalities that occur across schools.

For example, for gender-based inequalities, which our study shows primarily manifest within schools with substantial school variation in gender-based expected

wages, interventions should focus on school-specific practices. Schools could implement targeted professional development programs for teachers and counselors to address unconscious biases in advising students about CTE pathways (Threeton, 2007). Additionally, school-level mentorship programs pairing female students with successful women in high-wage CTE fields could be highly effective, particularly as complementary literature demonstrates that adult guidance and anticipated pay drives student's career anticipations (Aschbacher et al., 2010; Dick & Rallis, 1991; Wahl & Blackhurst, 2000). Marketing strategies challenging gender stereotypes in various CTE pathways could also be developed and implemented at the school level (LaCosse et al., 2020; Rainey et al., 2018). These within-school interventions are crucial because our results indicate significant variation in gender gaps across schools, suggesting that some schools have practices that successfully mitigate these gaps while others do not.

In contrast, our study reveals that racial and socioeconomic inequalities are largely driven by between-school differences, requiring a different approach. These school-level disparities mean that racial/ethnic and socioeconomic segregation will continue to generate inequality in expected wages unless schools serving minoritized and low-income students can offer POS of equal value. Policymakers could address this by improving the quality of CTE programs in schools where Black, Hispanic, and economically disadvantaged students are concentrated. While Delaware's school choice system could theoretically help students access higher-wage programs at other schools, Jacob and Ricks (2023) find that geographic distance between high-quality CTE programs and students' homes reduces program uptake, limiting school choice's effectiveness in addressing between-school inequalities.

Notably, we do not see systematic evidence that the availability of specific high wage pathways explains between school variability in expected wage gaps; rather, the group-level diversity across POS more consistently and strongly moderates selection into higher expected wage POS. What this means for policy is less

clear, though it suggests school culture may be a factor, causing disadvantaged students to focus on just one or two POS as opposed to encouraging student groups to identify those POS most suited to their own earnings potential.

To support these targeted interventions, we have provided a framework that allows for the feasible generation of a statewide data system that tracks CTE concentrators into their expected wages. This data-driven approach would allow for the identification of schools with successful practices in reducing specific gaps, facilitating knowledge sharing across the state.

4.8 Conclusion

CTE programs offer several benefits, including greater high school graduation rates, improved labor market participation, and increased earnings. However, there is substantial variation in these outcomes across different student subgroups (Brunner et al., 2021; Dougherty, 2018; Ecton & Dougherty, 2021; Hemelt et al., 2019). Using administrative data from the Delaware Department of Education and occupation employment and wage statistics from the Bureau of Labor Statistics, we estimate the expected wage gap between student subgroups.

Our estimation confirms that significant expected wage gaps exist across student subgroups. Female, Black, Hispanic, low-income, IEP participants, and ELL students are more likely to enroll in POS with lower expected wages, and these wage gaps vary substantially between schools. For gender, within-school factors contribute more to the wage gap between female and male students, while between-school factors are more significant for Black-White and Hispanic-White gaps. Our analysis on POS diversity suggests that within-school segregation in POS selection drives the wage gap, while school-level offerings of CTE programs of study do not moderate expected wage gaps.

Despite the importance of identifying the sources of wage gaps among CTE concentrators, limitations remain. First, we were unable to determine why students,

especially disadvantaged ones, are concentrated in specific POSs instead of pursuing a broader range of options. For instance, our data does not capture the influence of peer networks or peer effects, which may drive disadvantaged students toward limited POS options. Second, the results may not be generalizable (e.g., Jacob and Ricks (2023) emphasize school choice mechanisms as drivers of inequality). Third, we lack detailed qualitative information about the factors influencing students' decisions to concentrate in different POS and how those factors may differ across schools and by subgroup. Further exploration to collect alternative hypotheses, coupled with testable interventions, would allow policymakers to more carefully design policies to ensure more equitable engagement with the rich CTE offerings currently available to students.

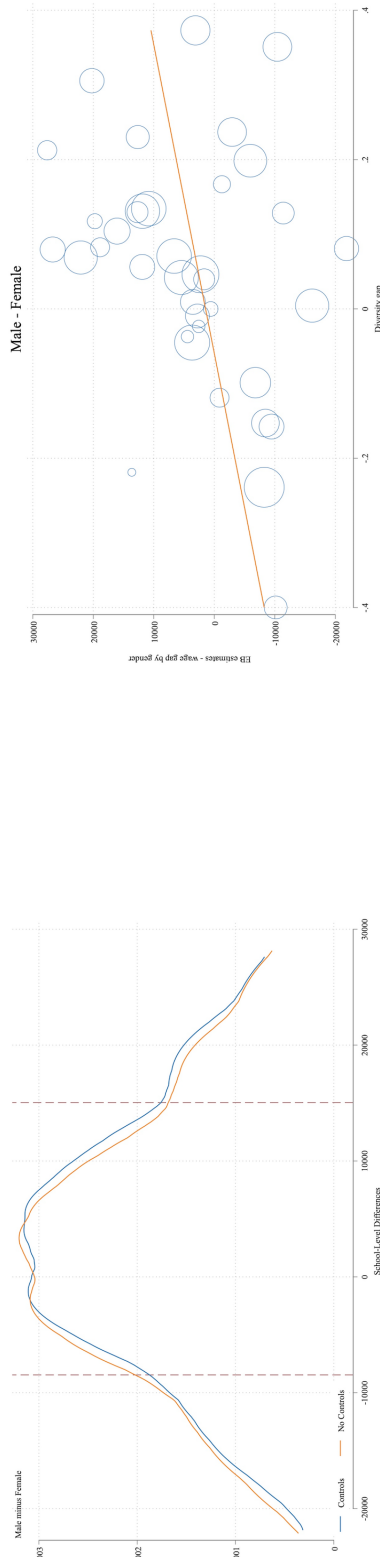
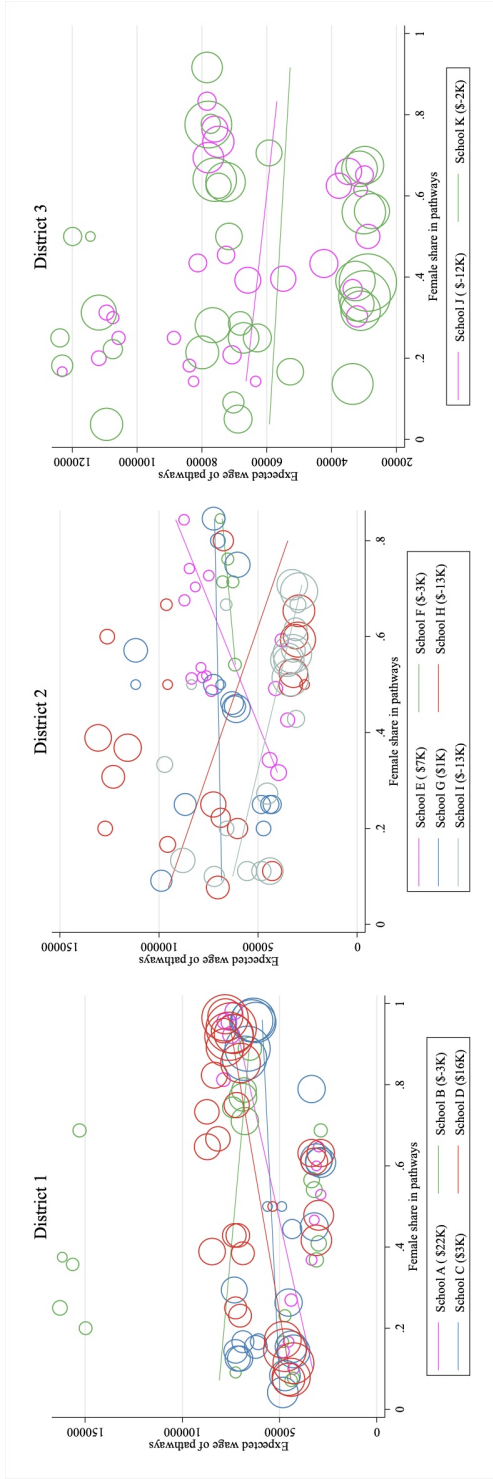


Figure 4.1: Gender wage gap distribution and female concentration in CTE pathways.

Notes: Panel A (top) shows the expected gender wage gaps and female proportions in POSs offered in schools from three public school districts in our dataset for 2017 – 2021 graduating cohort, anonymized here as District 1, District 2, and District 3. All school names have also been anonymized. Panel B (bottom left) is the distribution of the school-specific gender wage gap. The yellow line shows the estimated gender gap without test score adjustment and the blue line shows the gap with test score adjustment. Panel C (bottom right) shows the estimated gender gap at each school and the diversity gap which is calculated by subtracting the diversity index for females from the male diversity index. The scatterplot is weighted by using inverse variance weighting.

Table 4.1: Mean expected wages and student enrollment by career cluster

	State Total	C1	C2	C3	C4	C5	C6	C8	C9	C10	C11	C13	C14	C15	C16	C17
Mean expected wages	\$63K	\$39K	\$72K	\$68K	\$142K	\$78K	\$82K	\$74K	\$82K	\$98K	\$106K	\$61K	\$124K	\$103K	\$45K	\$80K
Student enrollment	18,329	5,073	706	1,201	433	473	483	3,333	2,594	35	156	108	701	1,868	907	258
Male	9,344	53%	84%	52%	51%	14%	67%	19%	51%	17%	87%	89%	63%	72%	86%	17%
(Female)	8,985	47%	16%	48%	49%	86%	33%	81%	49%	83%	13%	11%	37%	28%	14%	83%
FRPL	6,477	35%	26%	38%	24%	40%	33%	36%	46%	57%	19%	39%	35%	20%	39%	54%
(non-FRPL)	11,852	65%	74%	62%	76%	60%	67%	64%	54%	43%	81%	61%	65%	80%	61%	46%
IEP	2,172	13%	13%	15%	8%	12%	11%	15%	12%	3%	13%	19%	10%	11%	20%	31%
(non-IEP)	16,157	87%	87%	85%	92%	87%	89%	85%	88%	97%	87%	81%	90%	89%	80%	69%
ELL	2,472	13%	10%	15%	13%	13%	5%	7%	16%	3%	4%	9%	12%	7%	21%	9%
(non ELL)	15,857	87%	90%	85%	87%	86%	95%	93%	84%	97%	96%	91%	88%	93%	79%	91%
White	9,171	61%	57%	45%	52%	56%	48%	44%	32%	46%	62%	43%	47%	62%	46%	22%
Black	5,105	18%	20%	27%	31%	29%	32%	32%	48%	43%	20%	27%	35%	18%	25%	36%
Hispanic	2,990	15%	17%	19%	12%	12%	12%	19%	16%	6%	6%	24%	11%	11%	27%	38%
Asian	559	3%	4%	4%	3%	1%	6%	3%	1%	0%	10%	2%	4%	6%	1%	2%
Other race	504	3%	2%	4%	2%	3%	3%	2%	2%	6%	3%	5%	2%	3%	1%	2%

Notes: Expected wages are obtained from the occupational wage data of the U.S. Department of Labor, based on occupational codes assigned to each pathway. Dollar amounts are rounded to the nearest thousand for space. If more than one occupational code is assigned to a pathway, we use the mean of the assigned occupations' wages.

Mean expected wages are calculated based on the number of students in each pathway and the assigned occupational wages. Columns refer to the following career clusters: 1: Agriculture science, 2: Architecture, 3: Communication, Arts, & A/V Technology, 4: Business, 5: Education, 6: Finance, 8: Health Science, 9: Hospitality & Tourism, 10: Early Childhood & Cosmetology, 11: Information Technology, 13: Manufacturing, 14: Marketing, 15: STEM, 16: Automotive Technology, 17: Career Exploration.

Table 4.2: Regression results from multilevel mixed-effects models

	(1)	(2)	(3)	(4)
		w/o test score		w/ test score
	Mixed	Mixed (Random intercept)	Mixed	Mixed (Random intercept)
Panel A: Male				
Expected wage	2641.37*** (483.38)	2122.91*** (445.66)	3515.78*** (494.97)	2701.64*** (460.63)
Panel B: FRPL				
Expected wage	-7265.84*** (503.31)	-3196.75*** (481.36)	-4240.37*** (516.78)	-1726.60*** (488.43)
Panel C: IEP				
Expected wage	-7712.89*** (745.99)	-5145.40*** (684.83)	-658.17 (796.69)	-894.82 (737.04)
Panel D: ELL				
Expected wage	-2337.69*** (708.41)	857.15 (666.07)	-722.20 (702.51)	1187.32 (664.23)
Panel E: Race/Ethnicity				
Expected wage (Black)	-2659.05*** (569.66)	-451.54 (549.44)	1244.28** (591.89)	1545.64*** (563.76)
Expected wage (Hispanic)	-4948.63*** (687.56)	-918.45 (658.97)	-2032.78*** (693.62)	219.17 (662.25)
Expected wage (Asian)	10917.61*** (1420.72)	6231.74*** (1309.87)	8298.79*** (1405.98)	4851.58*** (1304.00)
Expected wage (Other)	-2098.40 (1491.94)	357.38 (1367.59)	-1003.35 (1470.59)	919.28 (1357.86)
N	18329	18329	18329	18329

Notes: Expected wages come from the U.S. Department of Labor occupational data, based on assigned occupation codes for each pathway. If more than one code is assigned, we use the mean wage. Panels show separate multilevel mixed-effects regressions for male vs. female, FRPL vs. non-FRPL, IEP vs. non-IEP, ELL vs. non-ELL, and white vs. other race/ethnicity. Columns (3) and (4) include controls for 8th grade English, Math, and Science test scores. All models include cohort-fixed effects. Standard errors are in parentheses. (***) $p < 0.01$, (**) $p < 0.05$, (*) $p < 0.1$

Table 4.3: Regression results from multilevel mixed-effects model

	(1)	(2)
	Mixed + Random Coefficient (w/o test score)	Mixed + Random Coefficient (w/ test score)
<i>Panel A: Male</i>		
Expected wage	3289.30 (2133.69)	3902.66* (2114.32)
SD of Random Slope	12509.34	12370.59
<i>Panel B: FRPL</i>		
Expected wage	-3200.22*** (1021.28)	-1652.37 (1042.25)
SD of Random Slope	5119.13	5239.20
<i>Panel C: IEP</i>		
Expected wage	-4751.73*** (1283.97)	-557.59 (1262.11)
SD of Random Slope	6055.14	5689.43
<i>Panel D: ELL</i>		
Expected wage	1061.24 (861.05)	1424.59 (872.24)
SD of Random Slope	2794.76	2909.91
<i>Panel E: Race/Ethnicity</i>		
Black: Expected wage	-584.14 (1117.76)	1494.54 (1110.13)
SD of Random Slope	5524.93	5422.92
Hispanic: Expected wage	-959.19 (897.27)	183.33 (892.76)
SD of Random Slope	3128.31	3074.89
Asian: Expected wage	7498.66*** (2481.50)	6264.87** (2446.58)
SD of Random Slope	29548.98	29320.44
Other: Expected wage	385.98 (1363.36)	942.34 (1353.66)
SD of Random Slope	5484.74	5380.63
N	18329	18329

Notes: Expected wages come from the U.S. Department of Labor occupational codes assigned to each pathway. If more than one code is assigned, we use the mean wage. Each panel represents a separate multilevel mixed-effects regression for male vs. female, FRPL vs. non-FRPL, IEP vs. non-IEP, ELL vs. non-ELL, and white vs. other race/ethnicity. Column (2) controls for 8th grade English, Math, and Science test scores. All models include cohort-fixed effects. Standard errors are in parentheses, and standard deviations of random slopes are presented below standard errors. (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Table 4.4: Decomposing Expected Wage Gap among Concentrators (w/ Test Scores)

	(1)	(2)	(3)	(4)
	Total Gap	Within-School	Between-School	Ambiguous
Panel A: Wage gap decomposition without test score adjustment				
Male – Female	2641.37	2187.22	301.27	71.50
(%)		83%	11%	3%
FRPL – non-FRPL	-7265.84	-2769.03	-3878.81	-270.11
(%)		38%	53%	4%
IEP – non-IEP	-7712.89	-4941.72	-2122.81	-108.11
(%)		64%	28%	1%
ELL – non-ELL	-2337.70	956.20	-3624.07	84.69
(%)		-41%	155%	-4%
Black – White	-2642.95	-526.90	-1742.30	-76.65
(%)		20%	66%	3%
Hispanic – White	-4979.18	-592.05	-4050.05	-105.80
(%)		12%	81%	2%
Asian – White	10976.08	5203.39	4564.18	229.77
(%)		47%	42%	2%
Panel B: Wage gap decomposition with test score adjustment				
Male – Female	3515.78	2983.92	409.25	122.61
(%)		85%	12%	3%
FRPL – non-FRPL	-4240.37	-1469.86	-2654.02	-116.49
(%)		35%	63%	3%
IEP – non-IEP	-658.17	219.13	-879.78	2.48
(%)		-33%	134%	0%
ELL – non-ELL	-722.20	1468.39	-2302.69	112.10
(%)		-203%	319%	-16%
Black – White	1611.88	2100.03	-776.40	288.26
(%)		130%	-48%	18%
Hispanic – White	-1882.75	686.03	-2680.35	111.57
(%)		-36%	142%	-6%
Asian – White	8136.98	3563.92	4394.17	178.89
(%)		44%	54%	2%

Notes: The wage gap decomposition follows Reardon (2008), where the student achievement gap is attributed to three parts: within-school, between-school, and ambiguous gap, based on Fryer and Levitt (2004) and Hanushek and Rivkin (2006). Within-school gap indicates the proportion of the wage gap attributable to within-school factors, while between-school gap captures the proportion explained by between-school factors. Ambiguous gap represents the portion where Fryer and Levitt (2004) attribute it to within-school factors, but Hanushek and Rivkin (2006) attribute it to between-school factors.

Table 4.5: Multilevel Mixed Effects Model with School-Level Characteristics

	(1) Male	(2) Black	(3) FRPL	(4) ELL	(5) IEP
Group = 1	2572.09*** (463.94)	1325.21** (538.90)	-1494.28*** (502.40)	542.86 (685.62)	-573.76 (759.43)
Mean expected wage of pathways (offered by school)	0.14*** (0.04)	0.28*** (0.04)	0.25*** (0.04)	0.27*** (0.04)	0.26*** (0.04)
Diversity gap (Group 1 - Group 0)	-7847.63*** (2984.87)	1907.12 (2891.13)	9333.91** (3891.75)	1999.73 (2029.56)	10499.18*** (2417.86)
Number of pathways offered	-223.12** (108.38)	-274.63** (107.97)	-224.49** (107.40)	-189.72* (105.44)	-191.43* (105.22)
Group = 1 × Mean expected wage of pathways	0.20*** (0.04)	-0.03 (0.05)	-0.01 (0.05)	-0.06 (0.06)	0.07 (0.06)
Group = 1 × Diversity gap	23911.42*** (2459.80)	30768.64*** (4521.68)	36469.52*** (6241.92)	21307.16*** (5668.95)	26262.89*** (6078.20)
Group = 1 × Number of pathways offered	81.73 (58.99)	34.16 (66.94)	-145.33** (64.51)	150.37* (81.90)	-191.78** (92.68)
Constant	63337.19*** (3402.22)	64122.03*** (3186.07)	64061.42*** (3115.42)	64562.85*** (3346.17)	65080.57*** (3326.51)
Random Effects Std. Dev. (group)	20352.30*** (2538.00)	19164.82*** (2382.63)	18808.12*** (2334.78)	20050.78*** (2505.15)	19922.98*** (2492.41)
Residual Std. Dev.	29322.37*** (153.35)	29423.47*** (153.97)	29405.14*** (153.83)	29460.53*** (154.42)	29453.71*** (154.36)
N	18323	18303	18314	18241	18246

Notes: Each column shows the expected wage gap between male vs. female, black vs. white, FRPL vs. non-FRPL, ELL vs. non-ELL, and IEP vs. non-IEP student groups. The mean expected wage of pathways follows the definition in Table 1. Number of pathways offered represents the number of pathways offered by the school the student attended. The diversity gap is calculated by subtracting the diversity indices of the reference group from that of the focus group. Standard Deviations of the Diversity gap of each group are: Male - Female (0.18), Black - White (0.12), FRPL (0.08), IEP (0.12), and ELL (0.14). (***) p<0.01, ** p<0.05, * p<0.1

Bibliography

- Abott, C., Kogan, V., Lavertu, S., & Peskowitz, Z. (2020). School district operational spending and student outcomes: Evidence from tax elections in seven states. *Journal of Public Economics*, 183, 104142.
- Advance CTE. (2023). Advancing the framework [Accessed: 2025-03-02].
- Altonji, J. G., Blom, E., & Meghir, C. (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annual Review of Economics*, 4(1), 185–223.
- Angrist, J., Bettinger, E., Bloom, E., King, E., & Kremer, M. (2002). Vouchers for private schooling in colombia: Evidence from a randomized natural experiment. *American economic review*, 92(5), 1535–1558.
- Aschbacher, P. R., Li, E., & Roth, E. J. (2010). Is science me? high school students' identities, participation and aspirations in science, engineering, and medicine. *Journal of Research in Science Teaching*, 47(5), 564–582.
- Association for Career and Technical Education. (2021). 2019-20 national cte enrollment data now available [Accessed: 2025-03-02]. <https://ctepolicywatch.acteonline.org/2021/05/2019-20-national-cte-enrollment-data-now-available.html>
- Baicker, K., & Gordon, N. E. (2004). *The effect of mandated state education spending on total local resources* (Working Paper No. 10701). National Bureau of Economic Research.

- Bailey, T., Jeong, D. W., & Cho, S.-W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review*, *29*(2), 255–270.
- Barnett, E. A., Kopko, E., Cullinan, D., & Belfield, C. R. (2020). *Who should take college-level courses? impact findings from an evaluation of a multiple measures assessment strategy* (tech. rep.). Center for the Analysis of Postsecondary Readiness.
- Bergman, P., Kopko, E., & Rodriguez, J. E. (2021). *Using predictive analytics to track students: Evidence from a seven-college experiment* (Working Paper No. w28948). National Bureau of Economic Research.
- Blackwell, M., Iacus, S., King, G., & Porro, G. (2009). Cem: Coarsened exact matching in Stata. *The Stata Journal*, *9*(4), 524–546.
- Blasnik, M. (2010). Reclink: Stata module to probabilistically match records.
- Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, *55*(3), 789–865.
- Boatman, A., & Long, B. T. (2018). Does remediation work for all students? how the effects of postsecondary remedial and developmental courses vary by level of academic preparation. *Educational Evaluation and Policy Analysis*, *40*(1), 29–58.
- Bobbitt-Zeher, D. (2007). The gender income gap and the role of education. *Sociology of Education*, *80*(1), 1–22.
- Brunner, E. J., Robbins, M. D., & Simonsen, B. (2018). Information, tax salience, and support for school bond referenda. *Public Budgeting & Finance*, *38*(4), 52–73.
- Brunner, E. J., Robbins, M. D., & Simonsen, B. (2021). Property tax information and support for school bond referenda: Experimental evidence. *Public Administration Review*, *81*(3), 488–499.

- Cabral, M., & Hoxby, C. (2012). *The hated property tax: Salience, tax rates, and tax revolts* (Working Paper No. 18514). National Bureau of Economic Research. <https://www.nber.org/papers/w18514>
- Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230.
- Candelaria, C. A., & Shores, K. A. (2019). Court-ordered finance reforms in the adequacy era: Heterogeneous causal effects and sensitivity. *Education Finance and Policy*, *14*(1), 31–60.
- Card, D., & Payne, A. A. (2002). School finance reform, the distribution of school spending, and the distribution of student test scores. *Journal of Public Economics*, *83*(1), 49–82. [https://doi.org/10.1016/S0047-2727\(00\)00177-8](https://doi.org/10.1016/S0047-2727(00)00177-8)
- Card, D., & Payne, A. A. (2021). High school choices and the gender gap in STEM. *Economic Inquiry*, *59*(1), 9–28.
- Carl D. Perkins Career and Technical Education Act of 2006. (2006). Pub. l. no. 115-224, 132 stat. 1433 (2018) [United States Federal Law].
- Cascio, E. U., Gordon, N., & Reber, S. (2013). Local responses to federal grants: Evidence from the introduction of title i in the south. *American Economic Journal: Economic Policy*, *5*(3), 126–159.
- Cattaneo, M. D., Jansson, M., & Ma, X. (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association*, *115*(531), 1449–1455.
- Cattaneo, M. D., Titiunik, R., Vazquez-Bare, G., & Keele, L. (2016). Interpreting regression discontinuity designs with multiple cutoffs. *The Journal of Politics*, *78*(4), 1229–1248.
- Cengiz, D., Dube, A., Lindner, A., & Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, *134*(3), 1405–1454.

- Chakrabarti, R., & Setren, E. (2011). *The impact of the great recession on school district finances: Evidence from new york* (Staff Report No. 534). Federal Reserve Bank of New York.
- Chen, J., & Roth, J. (2024). Logs with zeros? some problems and solutions. *The Quarterly Journal of Economics*, *139*(2), 891–936.
- Chen, X., Caves, L. R., Pretlow, J., Caperton, S. A., Bryan, M., & Cooney, D. (2020). *Courses taken, credits earned, and time to degree: A first look at the postsecondary transcripts of 2011-12 beginning postsecondary students. first look. nces 2020-501* (tech. rep.). National Center for Education Statistics.
- Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., & Cutler, D. (2016). The association between income and life expectancy in the United States, 2001–2014. *JAMA*, *315*(16), 1750–1766.
- Chingos, M. M., & Blagg, K. (2017). Do poor kids get their fair share of school funding. *Washington, DC: Urban Institute*, *3*.
- Cho, S.-W., Kopko, E., Jenkins, D., & Jaggars, S. S. (2012). *New evidence of success for community college remedial english students: Tracking the outcomes of students in the accelerated learning program (alp)* (CCRC Working Paper No. 53). Community College Research Center, Columbia University.
- Cho-Baker, S., Olivera-Aguilar, M., & Fishtein, D. (2021). Using latent class analysis to link career and technical education in adolescence and work and school transitions in young adulthood. *Career and Technical Education Research*, *46*(2), 59–79.
- Correia, S, Guimarães, P, & Zylkin, T. (2019). Ppmlhdf: Fast poisson estimation with high-dimensional fixed effects, arxiv. org.
- Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments. *American Journal of Sociology*, *106*(6), 1691–1730.
- Cullinane, J., & Treisman, P. U. (2010). *Improving developmental mathematics education in community colleges: A prospectus and early progress report on the*

- statway initiative* (NCPR Working Paper). National Center for Postsecondary Research.
- Darling-Hammond, L. (2019, August). *Investing for student success: Lessons from state school finance reforms* (Policy Brief). Learning Policy Institute. https://learningpolicyinstitute.org/sites/default/files/product-files/Investing_Student_Success_BRIEF.pdf
- Delaware Department of Education. (2017). CTE programs of study policies and procedures [Accessed: 2025-03-02].
- Delaware Department of Education. (2020). Career & technical education fiscal and accountability policies and procedures [Accessed: 2025-03-02].
- Denley, T. (2015). Co-requisite remediation pilot study—fall 2014 and spring 2015.
- Dick, T. P., & Rallis, S. F. (1991). Factors and influences on high school students' career choices. *Journal for Research in Mathematics Education*, 22(4), 281–292.
- Dixon-Román, E. J., Everson, H. T., & McArdle, J. J. (2013). Race, poverty and sat scores: Modeling the influences of family income on black and white high school students' sat performance. *Teachers College Record*, 115(4), 1–33.
- Domanico, R. (2024, February). Title i is a clunky, overbroad failure. low-income students deserve better.
- Domina, T., Penner, A., & Penner, E. (2017). Categorical inequality: Schools as sorting machines. *Annual Review of Sociology*, 43(1), 311–330.
- Dougherty, S. M. (2018). The effect of career and technical education on human capital accumulation: Causal evidence from massachusetts. *Education Finance and Policy*, 13(2), 119–148.
- Duncombe, W. D., Lukemeyer, A., & Yinger, J. Financing an adequate education: A case study of new york. In: In *Developments in school finance 2001–2002*. U.S. Department of Education, 2003, pp. 129–153. https://digitalscholarship.unlv.edu/sea_fac_articles/253

- Dynarski, S., Libassi, C., Michelmore, K., & Owen, S. (2018). *Closing the gap: The effect of a targeted, tuition-free promise on college choices of high-achieving, low-income students* (tech. rep.). National Bureau of Economic Research.
- Dynarski, S., Nurshatayeva, A., Page, L. C., & Scott-Clayton, J. (2022). *Addressing non-financial barriers to college access and success: Evidence and policy implications* (Working Paper No. w30054). National Bureau of Economic Research.
- Ecton, W. G., & Dougherty, S. M. (2021). *Heterogeneity in high school career and technical education outcomes* (EdWorkingPaper No. 21-492). Annenberg Institute at Brown University. <https://doi.org/10.26300/4jwf-wb39>
- Edgecombe, N. D., Cormier, M. S., Bickerstaff, S. E., & Barragan, M. (2013). Strengthening developmental education reforms: Evidence on implementation efforts from the scaling innovation project.
- Eggers, A. C., Freier, R., Grembi, V., & Nannicini, T. (2018). Regression discontinuity designs based on population thresholds: Pitfalls and solutions. *American Journal of Political Science*, *62*(1), 210–229.
- Evans, C. D., & Diekman, A. B. (2009). On motivated role selection: Gender beliefs, distant goals, and career interest. *Psychology of Women Quarterly*, *33*(2), 235–249.
- ExcelinEd. (2018). Perkins V reauthorization: Opportunities and challenges for states [Accessed: 2025-03-02].
- Fisher, R. C., & Papke, L. E. (2000). Local government responses to education grants. *National Tax Journal*, *53*(1), 153–168.
- Fryer Jr, R. G., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics*, *86*(2), 447–464.
- Gamoran, A. (1992). The variable effects of high school tracking. *American Sociological Review*, *57*(6), 812–828.

- Giani, M. S. (2019). Does vocational still imply tracking? examining the evolution of career and technical education curricular policy in Texas. *Educational Policy*, 33(7), 1002–1046.
- Gordon, N. (2004). Do federal grants boost school spending? evidence from title i. *Journal of Public Economics*, 88(9-10), 1771–1792.
- Gordon, N., & Reber, S. (2023, January). Title i of esea: How the formulas work. <https://all4ed.org/publication/title-i-of-esea-how-the-formulas-work/>
- Griffith, A. L., & Rask, K. N. (2014). Peer effects in higher education: A look at heterogeneous impacts. *Economics of Education Review*, 39, 65–77.
- Hainmueller, J., & Xu, Y. (2013). Ebalance: A Stata package for entropy balancing. *Journal of Statistical Software*, 54(7), 1–18.
- Hanushek, E. A., & Rivkin, S. G. (2006). *School quality and the Black-White achievement gap* (Working Paper No. w12651). National Bureau of Economic Research.
- Hemelt, S. W., Lenard, M. A., & Paepflow, C. G. (2019). Building bridges to life after high school: Contemporary career academies and student outcomes. *Economics of Education Review*, 68, 161–178.
- Hines Jr, J. R., & Thaler, R. H. (1995). Anomalies: The flypaper effect. *Journal of economic perspectives*, 9(4), 217–226.
- Hodara, M., & Xu, D. (2016). Does developmental education improve labor market outcomes? evidence from two states. *American Educational Research Journal*, 53(3), 781–813.
- Hoxby, C. M. (2001). All school finance equalizations are not created equal. *The Quarterly Journal of Economics*, 116(4), 1189–1231.
- Inman, R. P. (2008). *The flypaper effect* (tech. rep.). National Bureau of Economic Research.
- Jackson, C. K., Johnson, R. C., & Persico, C. (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms.

The Quarterly Journal of Economics, 131(1), 157–218. <https://doi.org/10.1093/qje/qjv036>

- Jacob, B. A., & Ricks, M. D. (2023). *Why choose career technical education? disentangling student preferences from program availability* (Working Paper No. w31756). National Bureau of Economic Research.
- Jaggars, S. S., & Bickerstaff, S. Developmental education: The evolution of research and reform (M. B. Paulsen, Ed.). In: *Higher education: Handbook of theory and research* (M. B. Paulsen, Ed.). Ed. by Paulsen, M. B. Cham: Springer, 2018, pp. 469–503.
- Kane, T. J., Boatman, A., Kozakowski, W., Bennett, C., Hitch, R., & Weisenfeld, D. (2019). *College remediation goes back to high school: Evidence from a statewide program in tennessee* (Working Paper No. w26133). National Bureau of Economic Research.
- Kane, T. J., Boatman, A., Kozakowski, W., Bennett, C., Hitch, R., & Weisenfeld, D. (2021). Is college remediation a barrier or a boost? evidence from the tennessee sails program. *Journal of Policy Analysis and Management*, 40(3), 883–913.
- Killingsworth, M. A. (2021). Experienced well-being rises with income, even above \$75,000 per year. *Proceedings of the National Academy of Sciences*, 118(4), e2016976118.
- Knight, D. S., & Mendoza, J. (2019). Does the measurement matter? assessing alternate approaches to measuring state school finance equity for california's local control funding formula. *AERA Open*, 5(3), 2332858419877424. <https://doi.org/10.1177/2332858419877424>
- Krueger, A. B. (2003). Economic considerations and class size. *The economic journal*, 113(485), F34–F63.

- LaCrosse, J., Canning, E. A., Bowman, N. A., Murphy, M. C., & Logel, C. (2020). A social-belonging intervention improves STEM outcomes for students who speak English as a second language. *Science Advances*, *6*(40), eabb6543.
- Lafortune, J., Rothstein, J., & Schanzenbach, D. W. (2018). School finance reform and the distribution of student achievement. *American Economic Journal: Applied Economics*, *10*(2), 1–26. <https://doi.org/10.1257/app.20160567>
- Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, *48*(2), 281–355.
- Lee, H., Shores, K., & Williams, E. (2022). The distribution of school resources in the united states: A comparative analysis across levels of governance, student subgroups, and educational resources. *Peabody Journal of Education*, *97*(4), 395–411.
- Logue, A. W., Douglas, D., & Watanabe-Rose, M. (2019). Corequisite mathematics remediation: Results over time and in different contexts. *Educational Evaluation and Policy Analysis*, *41*(3), 294–315.
- Logue, A. W., Watanabe-Rose, M., & Douglas, D. (2016). Should students assessed as needing remedial mathematics take college-level quantitative courses instead? a randomized controlled trial. *Educational Evaluation and Policy Analysis*, *38*(3), 578–598.
- Martorell, P., McFarlin, I. J., & Xue, Y. (2015). Does failing a placement exam discourage underprepared students from going to college? *Education Finance and Policy*, *10*(1), 46–80.
- Matsudaira, J. D., Hosek, A., & Walsh, E. (2012). An integrated assessment of the effects of title i on school behavior, resources, and student achievement. *Economics of Education Review*, *31*(3), 1–14.
- Mau, W.-C. J., & Li, J. (2018). Factors influencing STEM career aspirations of underrepresented high school students. *The Career Development Quarterly*, *66*(3), 246–258.

- McCall, L. (2001). Sources of racial wage inequality in metropolitan labor markets: Racial, ethnic, and gender differences. *American Sociological Review*, *66*(4), 520–541.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, *142*(2), 698–714.
- Meiselman, A. Y., & Schudde, L. (2022). The impact of corequisite math on community college student outcomes: Evidence from texas. *Education Finance and Policy*, 1–45. https://doi.org/10.1162/edfp.a_00373
- Melguizo, T., Ching, C. D., Ngo, F., & Harrington, D. (2021). AB 705 in the los angeles community college district: Results from fall 2019.
- Miller, T., Daugherty, L., Martorell, P., & Gerber, R. (2022). Assessing the effect of corequisite english instruction using a randomized controlled trial. *Journal of Research on Educational Effectiveness*, *15*(1), 78–102.
- Nagy, G., Garrett, J., Trautwein, U., Cortina, K. S., Baumert, J., & Eccles, J. S. Gendered high school course selection as a precursor of gendered careers: The mediating role of self-concept and intrinsic value (H. M. G. Watt & J. S. Eccles, Eds.). In: *Gender and occupational outcomes: Longitudinal assessments of individual, social, and cultural influences* (H. M. G. Watt & J. S. Eccles, Eds.). Ed. by Watt, H. M. G., & Eccles, J. S. Washington, DC: American Psychological Association, 2008, pp. 115–143.
- National Center for Education Statistics. (2023). Postsecondary institution expenses. condition of education. u.s. department of education, institute of education sciences [Retrieved March 02, 2025].
- National Center for Education Statistics. (2024). Public school expenditures. condition of education. u.s. department of education, institute of education sciences [Retrieved March 02, 2025].
- New America. (n.d.). Maintenance of effort [[Online; accessed 25-Feb-2025]].

- Ngo, F. (2019). Fractions in college: How basic math remediation impacts community college students. *Research in Higher Education*, *60*, 485–520.
- Nguyen-Hoang, P., & Yinger, J. (2020). The flypaper effect. *Journal of Education Finance*, *46*(2), 158–188.
- Oates, W. E. (1999). An essay on fiscal federalism. *Journal of Economic Literature*, *37*(3), 1120–1149. <https://doi.org/10.1257/jel.37.3.1120>
- of Education, D. D. (2023). Career and technical education in delaware, quick facts, funding and performance [Accessed: 2025-03-02]. https://www.doe.k12.de.us/cms/lib/DE01922744/Centricity/Domain/170/CTE%20in%20DE_Overview_.pdf
- O’Neill, J. E., & O’Neill, D. M. What do wage differentials tell about labor market discrimination? (S. W. Polachek, C. Chiswick, & H. Rapoport, Eds.). In: *The economics of immigration and social diversity* (S. W. Polachek, C. Chiswick, & H. Rapoport, Eds.). Ed. by Polachek, S. W., Chiswick, C., & Rapoport, H. Bingley, UK: Emerald Group Publishing, 2006, pp. 293–357.
- Rainey, K., Dancy, M., Mickelson, R., Stearns, E., & Moller, S. (2018). Race and gender differences in how sense of belonging influences decisions to major in STEM. *International Journal of STEM Education*, *5*, Article 10.
- Ran, F. X., & Lin, Y. (2022). The effects of corequisite remediation: Evidence from a statewide reform in tennessee. *Educational Evaluation and Policy Analysis*, *44*(3), 458–484. <https://doi.org/10.3102/01623737211070836>
- Reardon, S. F. (2008). Thirteen ways of looking at the black-white test score gap.
- Reardon, S. F., Kalogrides, D., & Shores, K. (2019). The geography of racial/ethnic test score gaps. *American Journal of Sociology*, *124*(4), 1164–1221.
- Reardon, S. F., Robinson-Cimpian, J. P., & Weathers, E. S. Patterns and trends in racial/ethnic and socioeconomic academic achievement gaps (H. F. Ladd & M. E. Goertz, Eds.). In: *Handbook of research in education finance and policy*

- (H. F. Ladd & M. E. Goertz, Eds.). Ed. by Ladd, H. F., & Goertz, M. E. New York, NY: Routledge, 2014, pp. 491–509.
- Reeves, R. V. (2013, September). Funding gaps in public schools: Real problem for social mobility, not parents' giving.
- Renzulli, L. A., Grant, L., & Kathuria, S. (2006). Race, gender, and the wage gap: Comparing faculty salaries in predominately White and historically Black colleges and universities. *Gender & Society, 20*(4), 491–510.
- Sacerdote, B. Peer effects in education: How might they work, how big are they and how much do we know thus far? In: In *Handbook of the economics of education*. Vol. 3. Elsevier, 2011, pp. 249–277.
- Sadler, P. M., Sonnert, G., Hazari, Z., & Tai, R. (2012). Stability and volatility of STEM career interest in high school: A gender study. *Science Education, 96*(3), 411–427.
- Scott-Clayton, J., Crosta, P. M., & Belfield, C. R. (2014). Improving the targeting of treatment: Evidence from college remediation. *Educational Evaluation and Policy Analysis, 36*(3), 371–393.
- Scott-Clayton, J., & Rodriguez, O. (2015). Development, discouragement, or diversion? new evidence on the effects of college remediation policy. *Education Finance and Policy, 10*(1), 4–45.
- Scrivener, S., Gupta, H., Weiss, M. J., Cohen, B., Cormier, M. S., & Brathwaite, J. (2018). Becoming college-ready: Early findings from a CUNY Start evaluation.
- Scrivener, S., Weiss, M. J., Ratledge, A., Rudd, T., Sommo, C., & Fresques, H. (2015). Doubling graduation rates: Three-year effects of cuny's accelerated study in associate programs (ASAP) for developmental education students.
- Shores, K., & Candelaria, C. (2020). Get real! inflation adjustments of educational finance data. *Educational Researcher, 49*(1), 71–74.

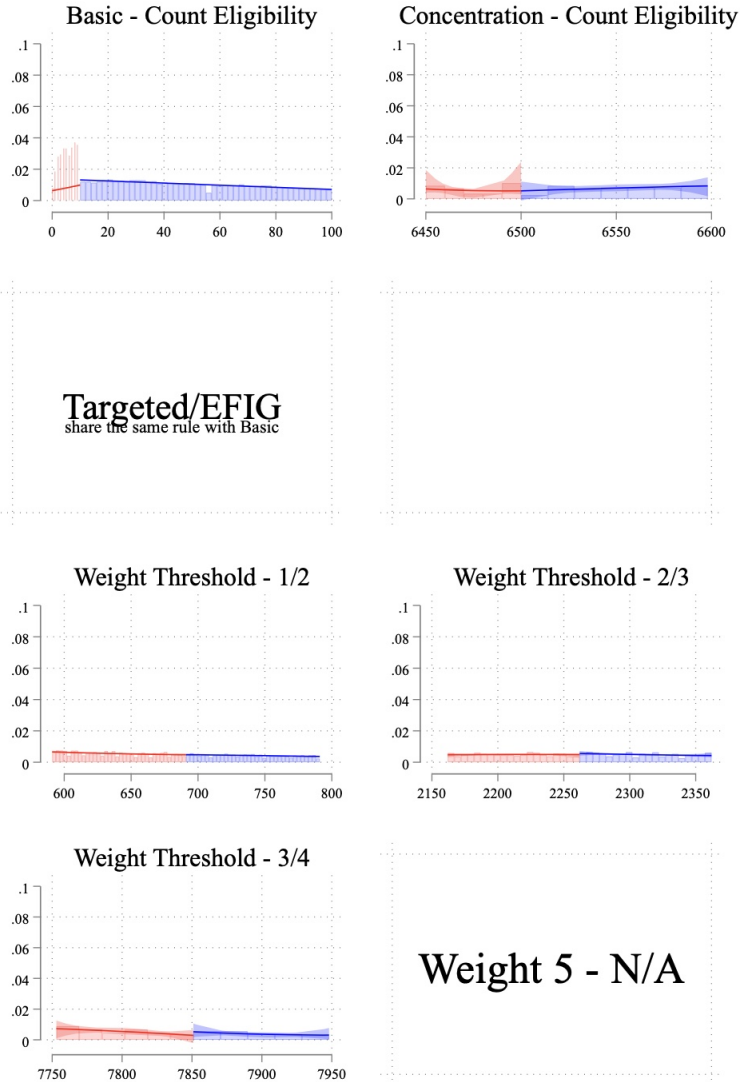
- Shores, K., Kim, H. E., & Still, M. (2020). Categorical inequality in Black and White: Linking disproportionality across multiple educational outcomes. *American Educational Research Journal*, 57(5), 2089–2131.
- Shores, K., Lee, H., & Williams, N. (2021, August). Expanding Title I could eliminate k–12 spending gaps—if the funds are well targeted.
- Signer, B., & Saldana, D. (2001). Educational and career aspirations of high school students and race, gender, class differences. *Race, Gender & Class*, 8(1), 22–34.
- Silva, J. S., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, 641–658.
- Skinner, R. R., & Riddle, W. (2020, November). *Esea: Title i-a poverty measures and grants to local education agencies and schools. CRS report R46600, version 2* (tech. rep. No. R46600) (ERIC Number: ED610713). Congressional Research Service. <https://eric.ed.gov/?id=ED610713>
- Sloane, C. M., Hurst, E. G., & Black, D. A. (2021). College majors, occupations, and the gender wage gap. *Journal of Economic Perspectives*, 35(4), 223–248.
- Steinberg, M. P., Quinn, R., Kreisman, D., & Anglum, J. C. (2016). Did pennsylvania’s statewide school finance reform increase education spending or provide tax relief? *National Tax Journal*, 69(3), 545–582.
- Stevens, A. H., Kurlaender, M., & Grosz, M. (2019). Career technical education and labor market outcomes: Evidence from California community colleges. *Journal of Human Resources*, 54(4), 986–1036.
- Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2), 175–199.
- Threeton, M. D. (2007). The Carl D. Perkins Career and Technical Education (CTE) Act of 2006 and the roles and responsibilities of CTE teachers and faculty members. *Journal of Industrial Teacher Education*, 44(1), 66–82.

- U.S. Department of Labor. (2024). Department of Labor equity action plan [Accessed: 2025-03-02].
- Van der Klaauw, W. (2008). Breaking the link between poverty and low student achievement: An evaluation of title i. *Journal of Econometrics*, *142*(2), 731–756.
- Verstegen, D. A. (2004). Towards a theory of adequacy: The continuing saga of equal educational opportunity in the context of state constitutional challenges to school finance systems. *Saint Louis University Public Law Review*, *23*(2). <https://scholarship.law.slu.edu/plr/vol23/iss2/5>
- Verstegen, D. A., & Jordan, T. S. (2009). A fifty-state survey of school finance policies and programs: An overview. *Journal of Education Finance*, 213–230.
- Wahl, K. H., & Blackhurst, A. E. (2000). Factors affecting the occupational and educational aspirations of children and adolescents. *Professional School Counseling*, *3*(5), 367–374.
- Winston, G., & Zimmerman, D. Peer effects in higher education (C. M. Hoxby, Ed.). In: *College choices: The economics of where to go, when to go, and how to pay for it* (C. M. Hoxby, Ed.). Ed. by Hoxby, C. M. Chicago: University of Chicago Press, 2004, pp. 395–424.
- Xu, D., & Dadgar, M. (2018). How effective are community college remedial math courses for students with the lowest math skills? *Community College Review*, *46*(1), 62–81.
- Xu, D., & Ran, F. X. (2022). The disciplinary differences in the characteristics and effects of non-tenure-track faculty. *Educational Evaluation and Policy Analysis*, *44*(1), 50–78.
- Zhao, K., Mokher, C. G., Park-Gaghan, T. J., & Hu, S. (2022). Who benefited more from the developmental education reform in florida? the role of exemption status. *AERA Open*, *8*, 23328584221126474.

Appendix A

PAPER ONE APPENDIX TABLES AND FIGURES

Figure A.1: Density of observations (count criteria)



Note: Each panel displays the distribution of the running variable around the indicated Title I grant/weight cutoff (vertical boundary). Red bars represent observations below the cutoff; blue bars are those above. The red/blue curves are local polynomial density estimates from `rddensity` using a polynomial order of 2 and a bias-correction order of 3, with a bandwidth of ± 100 (except for the Basic Grant, where the left bandwidth is 10). A visible jump at the cutoff indicates potential sorting or manipulation of observations.

Figure A.2: Density of Proportion of Eligible Kids, by State



Note: These panels show simple histograms of the re-centered running variable (all cutoffs normalized to 0) for each state, using the *proportion-based* stacked RDD dataset. The vertical red line marks the threshold at 0, allowing a direct visual check for any clustering around the cutoff.

Table A.1: McCrary Density Test (Count Sample)

	Weight	Transition	Cutoff	Obs Left	Obs Right	T-statistics	p-value	BW Left	BW Right
Basic	-		10	3126	35088	3.328	0.001	10	100
Concentration Targeted/EFIG			6500	19	40	-1.084	0.278	100	100
T1	1 → 1.5		691	3557	2759	-1.358	0.174	100	100
T2	1.5 → 2		2262	388	400	0.813	0.416	100	100
T3	2 → 2.5		7851	31	24	1.178	0.239	100	100
T4	2.5 → 3								

Notes: This table reports results from McCrary-type density tests (Cattaneo et al., 2020; McCrary, 2008) on the count-based sample. “Obs Left” and “Obs Right” denote the number of districts within the chosen bandwidth on each side of the cutoff. Failing to reject the null ($p > 0.05$) implies no significant manipulation at the threshold. “BW Left” and “BW Right” show the local polynomial bandwidths used on each side. All tests use a triangular kernel, polynomial order $p=2$, bias order $q=3$, and a chosen bandwidth of 100 (except BW left of Basic grant).

Table A.2: First Stage: Effects of Title I Formula Cutoffs on Title I Revenues

	(1)	(2)	(3)	(4)
	Per Eligible Title I			
	Basic		Non-Basic	
	FE Estimates	PML Estimates	FE Estimates	PML Estimates
Formula Cutoff	896.159*** (18.557)	2.335*** (0.426)	38.313*** (6.796)	0.028*** (0.005)
Proportion of Title I Kids	40.307 (55.777)	12.604*** (0.798)	20.054*** (2.210)	0.015*** (0.001)
Cutoff × Proportion	-40.522 (62.621)	-12.601*** (0.799)	-3.447 (2.884)	-0.002 (0.002)
Observations	7003	7003	65099	65099
Mean	805.0	805.0	1358.7	1358.7
R² (Pseudo R²)	0.354	0.380	0.537	0.472

Notes: This table reports estimates from a stacked regression discontinuity (RD) design using data from the 2008–09 through 2017–18 academic years. Columns (1) and (3) show results from an OLS model with fixed effects (“FE Estimates”), and Columns (2) and (4) use a Poisson pseudo–maximum likelihood estimation (“PML Estimates”) with Stata’s `ppmlhdfe`. All dollar amounts (per-eligible-child Title I) are adjusted to 2017 dollars. Both models include controls for district demographics (including racial composition, special education, and English learner proportions), as well as stack-by-state and stack-by-year fixed effects. Standard errors are clustered at the state level, with robust standard errors in parentheses. (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Table A.3: Effects on District Finance (Per Pupil, PML)

	(1)	(2)	(3)	(4)
<i>Panel A: Revenues</i>				
Revenues	Total Revenue	Federal Revenue	State Revenue	Local Revenue
Formula Cutoff	0.017 (0.013)	-0.005 (0.011)	0.010 (0.008)	0.021 (0.020)
N	72,102	72,102	72,102	72,102
Mean	15,342.8	1,332.1	7,038.4	6,972.3
Pseudo R²	0.370	0.609	0.460	0.439
<i>Panel B: Expenditures</i>				
Expenditures	Elem/Sec Exp	Capital Outlays	Services	Instruction
Formula Cutoff	0.012 (0.010)	-0.064* (0.034)	0.013 (0.010)	0.014 (0.011)
N	72,102	72,102	72,102	72,102
Mean	12,799.8	1,266.1	4,612.7	7,624.8
Pseudo R²	0.476	0.106	0.399	0.523

Notes: This table reports estimates from a stacked regression discontinuity (RD) design using data from the 2008–09 through 2017–18 academic years. All models use Poisson pseudo-maximum likelihood estimation, controlling for district demographics (including racial composition, special education, and English learner proportions), as well as stack-by-state and stack-by-year fixed effects. All dollar amounts are adjusted to 2017 school year dollars. Standard errors are clustered at the state level, with robust standard errors in parentheses. (* p<0.1, ** p<0.05, *** p<0.01)

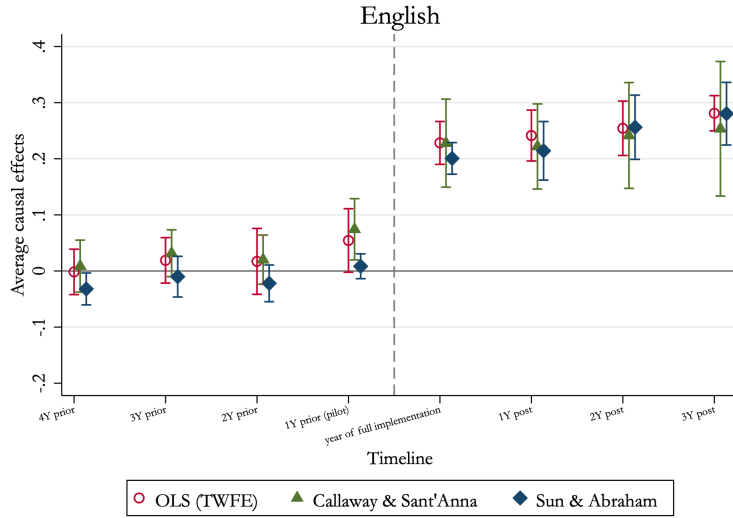
Table A.4: Effects on District Finance, Federal Source (Per Pupil)

Program	(1) IDEA	(2) Bilingual Education	(3) CNA
Formula Cutoff	9.291 (6.504)	0.481 (0.381)	-3.367* (1.812)
N	72,102	72,102	72,102
Mean	181.6	4.144	304.9
r²	0.177	0.092	0.534

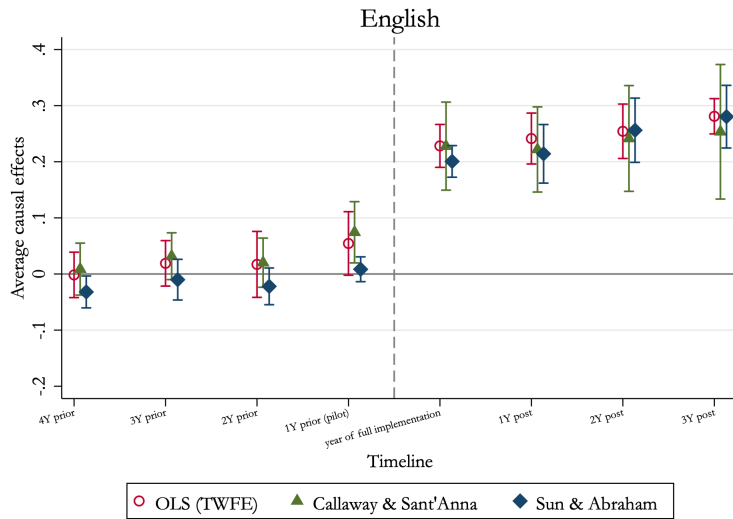
Notes: This table reports estimates from a stacked regression discontinuity (RD) design using data from the 2008–09 through 2017–18 academic years. Columns (1)–(3) show per-pupil federal revenues from three major programs: Individuals with Disability Education Act (IDEA), Bilingual Education, and Child Nutrition Act (CNA). All models use OLS with district-level demographic controls (racial composition, special education, English learner proportions) and stack-by-state and stack-by-year fixed effects, with standard errors clustered at the state level. All dollar amounts are in 2017 school year dollars. The modest negative coefficient on CNA partly explains the slight reduction in total federal revenues at Title I cutoffs, whereas IDEA and Bilingual funds remain unaffected. Standard errors are clustered at the state level, with robust standard errors in parentheses. (* p<0.1, ** p<0.05, *** p<0.01)

Appendix B

PAPER TWO APPENDIX TABLES AND FIGURES



(a) Math



(b) English

Figure B.1: Event study estimators comparison — effects on first-year gateway completion.

Note: These graphs show the point estimates and 95% confidence intervals for effects on first-year gateway completion rates from three models. Outcomes five years prior to the co-requisite reform were used as the reference for estimates in these plots. Estimates from two-way fixed-effects models are in red, heterogeneity-robust estimates using the method developed by Callaway & Sant'Anna (2021) are in green, and heterogeneity-robust estimates using methods by Sun & Abraham (2021) are in deep blue.

Table B.1: Evidence on the Effects of Traditional and Corequisite Developmental Education

Panel A. Treatment: traditional (prerequisite) developmental programs				
Authors	Method	Sample Size & Test Score Range	Data Source	Direction of Impacts on Key Outcomes
Bettinger & Long (2009)	IV	No restriction Obs: 28,376	Ohio Board of Regents	Persistence ↔, credential or transfer ↔ (↑)
Boatman & Long (2018)	Fuzzy RD	College-level vs. upper DE: BW: ±7 Obs: ~1,800 Upper vs. lower DE: BW: ±7 Obs: ~1,700	TBR, THEC	Complex effects: persistence ↓ or ↔; college credit ↔ or ↓
Bettinger & Long (2009)	IV	No restriction of test score range Observations: 28,376	Ohio Board of Regents	Persistence ↔, credential or transfer ↔ (↑)

(continued on next page)

Authors	Method	Sample Size & Test Score Range	Data Source	Direction of Impacts on Key Outcomes
Boatman & Long (2018)	Fuzzy RD	College-level vs. upper DE: Bandwidth: up to ±7 of college cutoff (COMPASS) Obs: ~ 1,800 Upper vs. lower DE: Bandwidth: up to ±7 of upper DE cutoff Obs: ~ 1,700	Tennessee Board of Regents (TBR), Tennessee Higher Education Commission (THEEC)	College-level vs. upper DE: Persistence ↓, grade in college-level course ↔, earned college-level credits & credential or transfer ↔ or ↓ Upper vs. lower DE: Persistence ↔, grade in college-level course ↔ (writing ↑), earned college-level credits & credential or transfer ↔ or ↓
Calcagno & Long (2008)	Fuzzy RD	Bandwidth: up to ±10 (Florida College Entry Level Placement Test) Obs: 98,370	Florida Dept. of Education	Persistence ↑, total credits ↑, college-level credits ↔, degree completion ↔

(continued on next page)

Authors	Method	Sample Size & Test Score Range	Data Source	Direction of Impacts on Key Outcomes
Clotfelter et al. (2015)	IV	No restriction of test score range Obs: ~ 14,000	North Carolina Education search Center	Short-run persistence ↔, enrollment & performance in college-level courses ↓, transfer & degree completion ↓
Dadgar (2012)	Fuzzy RD	Bandwidth: up to ±10 (COMPASS) Obs: 24,664	Virginia Community College System (VCCCS)	Passed college-level math ↔, earned credential ↓
De Paola & Scoppa (2014)	Fuzzy RD	Bandwidth: up to ±10 (U. of Calabria placement test) Obs: 4,019	University of Calabria	Earned credits by 2nd year ↑, drop out ↓
Martorell & McFarlin (2011)	Fuzzy RD	Bandwidth: ±10 (Texas Academic Skills Program) Obs: 255,878 (2-yr), 197,502 (4-yr)	Texas Schools Microdata Panel, THECB	Persistence & credential or transfer ↔

(continued on next page)

Authors	Method	Sample Size & Test Score Range	Data Source	Direction of Impacts on Key Outcomes
Melguizo et al. (2016)	RD	Bandwidth: optimal for each college, up to ± 14.9 (COMPASS or AC-CUPLACER) Obs: 16,553	Large urban community college district in California	Enroll & completion of college-level math courses \downarrow
Scott-Clayton & Rodriguez (2015)	RD	Bandwidth: ± 6 (COMPASS in-house exam) Obs: 100,250	Six colleges in large urban community college system	College enrollment \leftrightarrow , persistence \leftrightarrow (misplaced students \downarrow), pass college-level courses \downarrow , transfer/degree completion \leftrightarrow
Xu (2016)	Fuzzy RD	Bandwidth: up to ± 10 (COMPASS) Obs: 5,146	VCCS	Dropout \leftrightarrow , enroll & complete gatekeeper course \leftrightarrow , total credits earned in 5 yrs \leftrightarrow , total college-level credits in 5 yrs \leftrightarrow , transfer/degree completion \downarrow
Xu & Dadgar (2018)	Fuzzy RD	Bandwidth: ± 8 (COMPASS) Obs: 24,664	VCCS	Passed gatekeeper math \leftrightarrow , credential completion \downarrow

Panel B. Treatment: corequisite remediation

(continued on next page)

Authors	Method	Sample Size & Test Score Range	Data Source	Direction of Impacts on Key Outcomes
Boatman (2012)	Fuzzy RD	Bandwidth: ± 2 (ACT) Obs: 8,948	THEEC, TBR, IPEDS	Persistence \uparrow , credits attempted \uparrow
Boatman et al. (2021)	RD	Bandwidth: ± 11.95 (Duoc UC diagnostic test) Obs: 33,075	Duoc UC (Chile)	Dropout \downarrow , earned college-level credits by Y1 \leftrightarrow , first-term GPA \uparrow , grade in college algebra \uparrow
Duchini (2017)	RD	Bandwidth: up to ± 5 (unidentified test) Obs: 2,785	An anonymous university in north Italy	Persistence \leftrightarrow , passing college-level exam in remediation subjects \leftrightarrow , earned credits \leftrightarrow
Logue et al. (2016)	RCT	No restriction of test score range Obs: 717	Three universities in CUNY	Credits accumulated \uparrow , pass college-level statistics \uparrow
Logue et al. (2019)	RCT, PSM	No restriction of test score range Obs: 594	Three universities in CUNY	Quantitative course pass rates \uparrow , graduation rate \uparrow

(continued on next page)

Authors	Method	Sample Size & Test Score Range	Data Source	Direction of Impacts on Key Outcomes
Meiselman & Schudde (2022)	Fuzzy RD	Bandwidth: ± 5 (Texas Success Initiative test) Obs: 16,405	Texas ERC, THECB	Completion of math requirement \uparrow , degree completion \leftrightarrow
Miller et al. (2022)	RCT	No restriction of test score range Obs: 1,482	Five community colleges in Texas	Gateway English completion \uparrow , credit accumulation \uparrow , persistence \leftrightarrow
Ran & Lin (2022)	DiRD, RD	Bandwidth: ± 2 (ACT)	TBR	Gateway course completion \uparrow , enroll & complete subsequent college-level courses \uparrow , persistence & transfer \leftrightarrow

Notes: DID = difference in differences, DiRD = difference in regression discontinuity, IV = instrumental variable, PSM = propensity score matching, RCT = randomized control trial, RD = regression discontinuity, \uparrow = significant positive effects at 5% level, \downarrow = significant negative effects at 5% level, \leftrightarrow = insignificant results at 5% level.
Notes: DID = difference in differences, DiRD = difference in regression discontinuity, IV = instrumental variable, PSM = propensity score matching, RCT = randomized control trial, RD = regression discontinuity, \uparrow = significant positive effects at 5% level, \downarrow = significant negative effects at 5% level, \leftrightarrow = insignificant results at 5% level.

Table B.2: Summary statistics of different samples

	Raw data	FTIC sample	Recent HS graduate fall entrants sample
Panel A: All students			
Female	0.578	0.572	0.543
Age at college entry	23.995	19.544	18.486
White	0.713	0.715	0.705
Black	0.190	0.191	0.195
Hispanic	0.048	0.049	0.053
Other race	0.050	0.045	0.047
International Students	0.009	0.004	0.002
High school GPA	2.774	2.843	2.769
ACT Math	18.478	18.372	17.882
ACT English	20.100	19.706	18.645
ACT Reading	20.601	20.414	19.444
N	407,193	194,524	91,511
Panel B: Prerequisite students			
Female	0.583	0.580	0.551
Age at college entry	24.689	20.154	18.546
White	0.720	0.717	0.724
Black	0.201	0.205	0.198
Hispanic	0.036	0.039	0.039
Other race	0.043	0.039	0.039
International Students	0.009	0.005	0.003
High school GPA	2.871	2.906	2.873
ACT Math	18.226	18.080	17.793
ACT English	20.199	19.647	18.729
ACT Reading	20.534	20.265	19.398
N	231,238	109,747	42,904

(Continued on next page)

(Continued.) Summary statistics of different samples

	Raw data	FTIC sample	Recent HS graduate fall entrants sample
Panel C: Corequisite students			
Female	0.570	0.562	0.536
Age at college entry	23.084	18.755	18.432
White	0.705	0.711	0.687
Black	0.176	0.174	0.192
Hispanic	0.062	0.061	0.066
Other race	0.058	0.053	0.054
International Students	0.008	0.003	0.002
High school GPA	2.656	2.762	2.677
ACT Math	18.805	18.751	17.960
ACT English	19.969	19.783	18.570
ACT Reading	20.687	20.607	19.485
N	175,955	84,777	48,607

Notes: The raw data include any students who had enrollment records at TBR community colleges since 2010 fall. The FTIC sample includes first-time-in-college students who started at TBR between 2010 fall and 2018 fall, excluding those who were dual enrollment students during high school. The final analytic sample (recent HS graduate fall entrants) further excludes students who did not enter TBR community colleges in the fall semester of their high school graduation year. We imputed any missing data with school-cohort mean.

Table B.3: Covariate balance test

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Female	Black	Hispanic	Other race	No HS diploma	GED	Age at first term	HS GPA
Panel A. All Students								
Below cutoff	0.112*** (0.010)	0.099*** (0.016)	0.008* (0.003)	0.007** (0.002)	-0.003 (0.005)	0.002 (0.003)	0.175* (0.059)	-0.379*** (0.015)
Post coreq	-0.002 (0.003)	0.013* (0.004)	0.000 (0.005)	-0.009* (0.004)	-0.027* (0.011)	0.007 (0.004)	0.050 (0.034)	-0.030 (0.025)
Below cutoff \times post coreq	-0.006 (0.009)	-0.012* (0.003)	0.006 (0.005)	-0.004 (0.003)	0.005 (0.006)	-0.002 (0.003)	-0.074 (0.047)	-0.011 (0.015)
Panel B. Test score subgroup								
Group 1 \times post coreq	-0.010 (0.009)	-0.031** (0.009)	0.004 (0.007)	-0.010* (0.003)	0.011 (0.010)	-0.003 (0.004)	-0.178 (0.100)	-0.049 (0.027)
Group2 \times post coreq	-0.009 (0.010)	-0.017* (0.006)	0.005 (0.005)	-0.002 (0.003)	0.008 (0.007)	-0.003 (0.003)	-0.084 (0.060)	-0.006 (0.020)
Group3 \times post coreq	0.005 (0.013)	0.001 (0.007)	0.007 (0.004)	-0.003 (0.005)	-0.002 (0.004)	-0.000 (0.001)	-0.005 (0.016)	0.021 (0.013)
N	91,511	91,511	91,511	91,511	91,511	91,511	91,511	91,511

Notes: Estimates use the student covariate in each column as the dependent variable, controlling for college fixed effects, cohort fixed effects, and high school fixed effects. We used the difference-in-differences specifications without other student covariates in the model to provide more conservative results. Standard errors are clustered at the college-, cohort-, and high school-level; robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table B.4: Effects of corequisite remediation on first-year gateway course completion: Robustness check (excluding data from academic year 2019–20)

	Two-way fixed-effects estimates		Heterogeneity-robust estimates	
	(1)	(2)	(3)	(4)
Panel A. DID estimates				
	Math	English	Math	English
Below cutoff * coreq	0.205*** (0.019)	0.223*** (0.031)	0.193*** (0.039)	0.238*** (0.039)
Baseline mean (Baseline SD)	0.337 (0.472)	0.604 (0.489)	0.337 (0.472)	0.604 (0.489)
N	80,447	80,447	304	304
Panel B. Event study estimates				
	Two-way fixed-effects estimates		Heterogeneity-robust estimates	
	Math	English	Math	English
Pre4 * below cutoff	0.002 (0.013)	0.003 (0.016)	-0.027 (0.026)	-0.030* (0.015)
Pre3 * below cutoff	0.003 (0.016)	0.023 (0.018)	-0.014 (0.021)	-0.009 (0.019)
Pre2 * below cutoff	0.002 (0.019)	0.024 (0.034)	-0.005 (0.027)	-0.021 (0.016)
Pre1 * below cutoff	0.098*** (0.014)	0.067 (0.031)	0.087** (0.023)	0.010 (0.011)
Pre0 * below cutoff	0.239*** (0.020)	0.234*** (0.016)	0.227*** (0.026)	0.199*** (0.015)
Post1 * below cutoff	0.235*** (0.022)	0.247*** (0.030)	0.229*** (0.029)	0.216*** (0.025)
Post2 * below cutoff	0.205*** (0.012)	0.256*** (0.022)	0.224*** (0.024)	0.262*** (0.033)
N	80,447	80,447	80,447	80,447

Notes: Results are based on analyses for students from 2010–11 to 2017–18 cohorts. Gateway course outcomes are tracked up to one academic year since the initial term enrolled at TBR. The TWFE models controlled for students' demographic and pre-college academic characteristics (Table 1 Panel A), plus college, cohort, and high school fixed-effects. The heterogeneity-robust DID model was conducted at the college-cohort-remedial status level. Standard errors are clustered at college, cohort, and high school levels. Robust standard errors in parentheses. Pre-periods stand for years before the corequisite implementation; post-periods for years after. Pre0 represents the year of corequisite reform. We used pre5 as the reference group for all event study specifications. *** p<0.01, ** p<0.05, * p<0.1

Table B.5: Effects of corequisite remediation on credit accumulation, persistence, transfer, and completion by end of Y3: Robustness check (excluding data from academic year 2019–20)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	
	Total number of credits			College-level credits		Enrollment		Credential
	Enrolled	Earned	Enrolled	Earned	Left TN college system	Public Transfer to 4yr inst.	Earned any credential	
Panel A: All Students								
Below cutoff * coreq	-1.346** (0.349)	-1.398** (0.302)	2.474* (0.722)	0.804 (0.412)	0.041** (0.008)	-0.014** (0.003)	-0.025** (0.006)	
Baseline mean & SD	44.587 (23.044)	33.001 (25.417)	39.251 (24.057)	29.773 (25.080)	0.560 (0.496)	0.097 (0.296)	0.122 (0.328)	
N	68,091	68,091	68,091	68,091	68,091	68,091	68,091	
Panel B: Subgroup Analysis								
Group 1 * coreq	-1.402 (0.870)	-1.271 (0.784)	4.685** (1.329)	1.855 (0.911)	0.044* (0.014)	-0.024** (0.006)	-0.054** (0.011)	
Group 2 * coreq	-1.390* (0.419)	-1.345** (0.384)	2.604** (0.631)	0.928* (0.316)	0.045** (0.011)	-0.016* (0.005)	-0.027** (0.005)	
Group 3 * coreq	-1.354** (0.316)	-1.679*** (0.281)	0.685 (0.442)	-0.247 (0.353)	0.034*** (0.004)	-0.006 (0.005)	-0.008 (0.006)	
Baseline mean & SD	36.396 (21.732)	21.840 (22.208)	25.060 (21.179)	15.996 (19.638)	0.697 (0.460)	0.039 (0.194)	0.027 (0.163)	
Group 1	42.733 (22.987)	30.154 (24.626)	35.327 (23.187)	25.528 (23.364)	0.591 (0.492)	0.072 (0.258)	0.076 (0.264)	
Group 2	47.159 (22.927)	36.440 (25.544)	43.416 (23.050)	33.777 (24.959)	0.523 (0.499)	0.110 (0.312)	0.143 (0.350)	
Group 3	68,091	68,091	68,091	68,091	68,091	68,091	68,091	

Notes: Results are based on analyses for students from 2010–11 to 2016–17 cohorts. All outcomes are tracked up to three years since the initial term enrolled at TBR. All models controlled for students' demographic and pre-college academic characteristics (see Table 1 Panel A), college fixed effects, cohort fixed effects, and high school fixed effects. Standard errors are clustered at the college-, cohort-, and high school-level. Robust standard errors in parentheses. Group 1 (lowest scoring DE group) = students whose ACT scores were 6+ points below cutoff; Group 2 = 3–5 points below; Group 3 = 1–2 points below. Group 4 (the reference group) has scores at or above the cutoff. We used students' minimum test scores of writing and reading to define the test score subgroup for English.

*** p<0.01, ** p<0.05, * p<0.1

Table B.6: Effects of corequisite remediation on gateway course completion and downstream outcomes: Robustness check (excluding students from high schools with SAILS program)

Outcomes	Below cutoff * post coreq
Complete gateway math by year one	0.236*** (0.018)
Complete gateway English by year one	0.243*** (0.033)
Number of credits enrolled by Y3	-0.369 (0.505)
Number of credits earned by Y3	-0.504 (0.480)
Number of college-level credits enrolled by Y3	3.330** (0.803)
Number of college-level credits earned by Y3	1.643* (0.558)
Left TN community colleges & public universities by Y3	0.027* (0.012)
Transfer to a four-year university by Y3	-0.021 (0.012)
Earned any credential by Y3	-0.020* (0.006)
N	59,686

Notes: Results are based on students entering TBR between 2010-11 to 2017-18 academic years, excluding students graduated from high school implemented SAILS program during their senior year. The sample restriction represented 1% of students of 2013-14 cohort, 16% of 2014-15 cohort, 34% of 2015-16 cohort, 51% of 2016-17 cohort, and 57% of 2017-18 cohort. Each row presents the DID estimates of corequisite remediation effect on a separate outcome. All models controlled for students' demographic and pre-college academic characteristics shown in Table 1 Panel A, college fixed-effects, cohort fixed-effects, and high school fixed-effects. Standard errors are clustered at college-, cohort-, and high school-level. Robust standard errors are shown in parentheses. (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Appendix C

PAPER THREE TECHNICAL APPENDIX: MEDEL SENSITIVITY CHECK

To ensure that our multilevel mixed-effects model results are robust to model selection or selection of covariate balancing methods, we employed multiple alternative approaches and compared the results from these alternative methods with our main results. Specifically, we used Ordinary Least Squares (OLS), Coarsened Exact Matching (CEM, Blackwell et al., 2009), and Entropic Balancing (`ebalance`) methods (Hainmueller & Xu, 2013). First, we used OLS methods to see if our preferred model results deviated significantly from the simplest regression model. The regression model can be written as follows:

$$Y_{ij} = \beta_0 + \beta_1 D_{ij} + X_{ij} + \epsilon_{ij} \quad (\text{C.1})$$

Here, Y_{ij} represents the mean expected wages of pathway that student i in school j enrolled in. D_{ij} is binary variables that shows students demographics (gender, race/ethnicity, FRPL status, IEP status, and ELL status). X_{ij} is the vector of covariates, which includes students ELA, math, science test scores and graduating cohort. ϵ_{ij} is student-specific standard errors. We are interested in estimating β_1 , the estimated wage gap between student subgroups (i.e., male vs female, white vs other racial groups).

Next, we use CEM methods to improve covariate balance. Based on matched sets created from CEM, we estimated weighted least square model and compare these results with our main model. CEM is the matching methods that can be useful

for improving the match between treatment and control group. Exact matching methods matches treatment and control group observations only when they have the same value in the covariates of the interests. This will create only a small number of matches, especially when covariates of interests are continuous. Instead, CEM coarsens (groups) the data into intervals and generates strata based on the bins created from the coarsened covariates. Observations in the same stratum function as a matched set. In our data, we create matches based on students' 8th grade ELA, math and science test scores and high school graduating cohort. We divide each test subjects into 6 bins and create a bin for each graduating cohort for every public high school in our sample. When strata are generated, the weights are also assigned to the stratum if the stratum is matched. After creating strata and weights, we run the following regression model for each school (j):

$$Y_i = \beta_0 + \beta_1 D_i + \epsilon_i \tag{C.2}$$

Y_i is the mean of expected wage from the pathway that student i enrolled in, and D_i represents binary variables of our interests, which includes gender, race/ethnicity, ELL, FRPL, and IEP status. Here, we do not include covariate in Equation (A.2) since CEM already controls for the covariates when creating strata. This regression is weighted using the CEM weights. Again, we are interested in measuring β_1 , the estimated expected wage gap between student subgroups.

Another approach we use to improve covariate balance is entropic balancing (ebalance). Ebalance creates weights that ensure covariate balance between the treatment and control groups by generating weights w_{ij} that minimize entropy distance $\sum_{i \in R, i \in j} w_{ij} \log w_{ij}$ with respect to $\sum_{i \in R, i \in j} w_{ij} x_{ijk} = \sum_{i \in F, i \in j} x_{ijk}, \forall k$ for student i in school j . Here, the subscript F denotes the focus group (male, black or Hispanic, ELL, FRPL and IEP students) and R means reference group (female, white, non-ELL, non-FRPL and non-IEP students). k represents covariate of our

interest, which includes 8th grade test scores and high school graduating cohort. Note that in our analysis, we are not looking for causal impacts of a treatment, but estimating the gap estimates between student demographics, so we use focus group and reference group instead of treatment and control groups.

After generating weights from the covariates, we created a weighted dataset for each demographic variable and then ran a simple OLS model with the reweighted datasets. The OLS model for school j can be written as follows:

$$Y_i = \beta_0 + \beta_1 D_i + \epsilon_i \tag{C.3}$$

This model equation is equivalent to Equation (A.2). However, it differs in estimation, as Equation (A.3) is weighted using ebalance weights. The model is also different from Equation (A.1) since it does not include covariates since the weighted datasets already controlled covariates by reweighting.

After estimating models (A.1) – (A.3), we compared these results with our main analysis. Firstly, we obtained the correlation coefficients of the estimated β_1 s from our main model (1) and alternative methods (A.1 – A.3). The correlation coefficients are presented in Table A1, Panel A. The baseline model is the multilevel mixed-effects model without test score adjustment. For gender, the estimated wage gaps show the highest correlation across models, all above 0.98. The coefficients from the models for Black, Asian, FRPL, and IEP are slightly lower than those for gender. The weakest correlations between the multilevel mixed-effects model and the alternative methods are found in the CEM Hispanic and CEM ELL groups. However, these correlations are still strong enough to conclude that our multilevel mixed-effects model results are robust to alternative model selection and covariate balancing methods.

Next, we examined the estimated standard deviations of the random effects or the standard deviations of the school-specific estimated wage gaps by methods.

These results are presented in Table [D.2](#), Panel B. The mixed-effects models and the ebalance method show similar magnitudes of standard deviations of wage gaps, while the OLS models and CEM show much smaller standard deviations. Although the standard deviations from the OLS models and CEMs are smaller, the trends shown by the groups are quite similar. Therefore, our main results are not solely the product of the multilevel model we employed but reflect the actual phenomena of Delaware career technical education.

Appendix D

PAPER THREE APPENDIX TABLES AND FIGURES

Table D.1: Regression results from fixed-effects model

	(1)	(2)	(3)	(4)
	w/o test score		w/ test score	
	Cohort FE	Cohort & School FE	Cohort FE	Cohort & School FE
Panel A: Male				
Expected wage	2641.37*** (483.46)	2118.91*** (445.77)	3515.78*** (495.09)	2698.17*** (460.78)
Panel B: Low-income				
Expected wage	-7265.84*** (503.39)	-3171.93*** (481.51)	-4240.37*** (516.91)	-1708.33*** (488.58)
Panel C: Special ED				
Expected wage	-7712.89*** (746.11)	-5124.00*** (685.03)	-658.17 (796.89)	-880.43 (737.37)
Panel D: ELL				
Expected wage	-2337.69*** (708.53)	867.50 (666.22)	-722.20 (702.69)	1193.18* (664.43)
Panel E: Race/Ethnicity				
Expected wage (Black)	-2659.05*** (569.80)	-430.58 (549.71)	1244.28** (592.09)	1558.05*** (564.04)
Expected wage (Hispanic)	-4948.63*** (687.73)	-896.87 (659.21)	-2032.78*** (693.85)	233.46 (662.52)
Expected wage (Asian)	10917.61*** (1421.07)	6206.60*** (1310.21)	8298.79*** (1406.44)	4829.66*** (1304.44)
Expected wage (Other)	-2098.40 (1492.31)	355.74 (1367.90)	-1003.35 (1471.08)	915.24 (1358.26)
N	18329	18329	18329	18329

Notes: Expected wages come from U.S. Department of Labor occupational data based on assigned occupational codes. If more than one code is assigned, we use the mean wage. Each panel shows results from separate regression models for male vs. female, low-income vs. non-low-income, special ED vs. non-special ED, ELL vs. non-ELL, and white vs. other race/ethnicity. Columns (3) and (4) include 8th grade English, Math, and Science test scores as controls. All models include cohort fixed effects; columns (2) and (4) also include school fixed effects. Standard errors are in parentheses. (** * $p < 0.01$, * $p < 0.05$, * $p < 0.1$)

Table D.2: Correlation & SD of Anticipated Wage Gaps (Weighted Coefficients)

	Gender	Black	Hispanic	Asian	FRPL	IEP	ELL
<i>Panel A: Correlations of Anticipated Wage Gaps between Unadjusted Random Effects and Alternative Models</i>							
Mixed (w/ test score)	0.999	0.992	0.988	0.997	0.991	0.996	0.989
OLS (w/o test score)	0.997	0.984	0.853	0.948	0.864	0.936	0.843
OLS (w/ test score)	0.991	0.903	0.838	0.920	0.813	0.811	0.867
CEM	0.988	0.853	0.776	0.850	0.829	0.825	0.793
EBALANCE	0.991	0.890	0.879	0.931	0.915	0.848	0.881
<i>Panel B: Standard Deviation of Random Effects or School-Specific Anticipated Wage Gaps</i>							
Mixed (w/o test score)	11812	4616	1841	10280	3927	4759	1564
Mixed (w/ test score)	11717	4558	1811	10104	4040	4398	1700
OLS (w/o test score)	6247	5710	2624	8657	3053	3824	2633
OLS (w/ test score)	6659	5480	2461	7813	2517	2776	2735
CEM	3486	2802	1754	1287	1625	2021	1379
EBALANCE	11541	4537	3773	4976	3264	5007	3683
N	36	35	36	31	37	35	36

Notes: Panel A shows correlation coefficients of expected wage gaps between the test-score unadjusted random effects model and alternative models. CEM represents the coarsened exact matching method (Blackwell et al., 2009), and EBALANCE represents the entropy balancing method (Haimueller & Xu, 2013). Correlation coefficients are obtained using inverse variance weighting. Panel B shows the standard deviations of random effects or school-specific anticipated wage gaps estimated from multilevel mixed-effects models and the alternative models.

Table D.3: Multilevel Mixed Effects Model with School-Level Characteristics

	(1) Male	(2) Black	(3) FRPL	(4) ELL	(5) IEP
Mean Expected Wages					
Group = 1	2693.22*** (468.37)	1140.05** (542.31)	-1502.08*** (502.19)	503.30 (685.27)	-623.53 (769.90)
Mean expected wage of pathways (offered by school)	0.14*** (0.04)	0.31*** (0.04)	0.24*** (0.04)	0.23*** (0.04)	0.29*** (0.04)
Diversity gap (Group 1 - Group 0)	-8081.34*** (3041.13)	1850.34 (2947.75)	251.83 (4833.36)	970.81 (2038.96)	10441.14*** (2506.08)
Number of pathways offered	-215.92*** (109.19)	-287.95*** (109.78)	-214.03* (111.02)	-115.43 (108.49)	-250.43** (105.76)
Diversity gap (Group 1 - Group 0) × Number of pathways offered	-16.13 (432.68)	-1348.91** (591.92)	-1532.79** (660.63)	-1297.07*** (453.38)	1533.53*** (393.42)
Group=1 × Mean expected wage of pathways	0.19*** (0.04)	-0.05 (0.05)	-0.01 (0.05)	-0.07 (0.06)	0.06 (0.06)
Group = 1 × Diversity gap	22775.87*** (2539.96)	29734.63*** (4556.39)	47167.07*** (6748.08)	21738.41*** (5705.12)	26459.62*** (6141.97)
Group = 1 × Number of pathways offered	89.90 (59.18)	40.10 (66.96)	-149.24** (64.50)	115.07 (104.48)	-194.53** (92.65)
Group = 1 × Diversity gap × Number of pathways offered	-770.99* (432.89)	2476.38*** (842.45)	3370.49*** (924.54)	594.29 (1049.45)	647.88 (1126.96)
Diversity gap (Group 1 - Group 0) × Mean expected wage of pathways	-0.20 (0.16)	0.22 (0.23)	0.14 (0.26)	0.68*** (0.20)	1.07*** (0.20)
_cons	63428.98*** (3404.16)	64071.28*** (3077.74)	63993.01*** (3039.23)	64329.26*** (3448.58)	64745.59*** (3289.74)
Random Effects Std. Dev. (group)	20341.96*** (2544.40)	18465.55*** (2339.41)	18309.48*** (2319.48)	20674.78*** (2585.23)	19692.61*** (2465.58)
Residual Std. Dev.	29318.16*** (153.33)	29415.48*** (153.93)	29392.78*** (153.77)	29438.86*** (154.31)	29427.30*** (154.22)
N	18323	18303	18314	18241	18246

Notes: Each column shows the expected wage gap between male vs. female, black vs. white, FRPL vs. non-FRPL, ELL vs. non-ELL, and IEP vs. non-IEP student groups. The mean expected wage of pathways follows the definition in Table 1. "Number of pathways offered" represents the number of pathways offered by the school that the student attended. The diversity gap is calculated by subtracting the diversity indices of the reference group from that of the focus group. Standard Deviations of the Diversity gap of each group are: Male-Female (0.18), Black-White (0.12), FRPL (0.08), IEP (0.12), ELL (0.14). (** * $p < 0.01$, * $p < 0.05$, * $p < 0.1$)

Table D.4: Multilevel Mixed Effects Model with School-Level Characteristics

	(1) Male	(2) Black	(3) FRPL	(4) ELL	(5) IEP
Mean Expected Wages					
Group = 1	2612.15*** (469.86)	1162.12** (544.53)	-1492.71*** (502.96)	667.82 (694.46)	-671.94 (770.48)
Mean expected wage of pathways	0.17*** (0.04)	0.31*** (0.04)	0.25*** (0.04)	0.26*** (0.04)	0.26*** (0.04)
Diversity gap (Group 1 - Group 0)	-7754.38** (3057.28)	1507.03 (2906.98)	292.36 (4808.03)	1517.01 (2033.94)	11751.09*** (2501.47)
Number of pathways offered	-231.59** (108.39)	-268.33** (108.78)	-211.75* (110.91)	-103.73 (108.21)	-217.47** (105.65)
Diversity gap × Num. of pathways	167.73 (436.30)	-1843.00*** (584.53)	-1709.01*** (548.35)	-1579.50*** (444.80)	730.41** (364.59)
Num. of pathways × Mean exp. wage	0.00 (0.00)	-0.01* (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
Group=1 × Mean exp. wage	0.16*** (0.04)	-0.04 (0.05)	-0.00 (0.05)	-0.05 (0.06)	0.07 (0.06)
Group = 1 × Diversity gap	21818.71*** (2592.28)	29926.13*** (4563.44)	47056.02*** (6742.23)	21494.21*** (5706.94)	26149.29*** (6146.76)
Group = 1 × Number of pathways	102.94* (59.47)	37.54 (67.16)	-151.58** (64.91)	124.94 (105.11)	-198.14** (92.92)
Group 1 × Div. gap × Num. of pathways	-1049.34** (460.15)	2476.62*** (844.50)	3384.62*** (934.95)	671.35 (1051.85)	783.49 (1138.04)
Group 1 × Num. of paths × Mean exp. wage	-0.01* (0.01)	0.00 (0.01)	0.00 (0.01)	0.01 (0.01)	-0.01 (0.01)
_cons	63300.53*** (3335.92)	63941.15*** (2959.76)	63948.93*** (2987.93)	64103.82*** (3323.93)	64917.74*** (3289.11)
Random Effects Std. Dev. (group)	19907.48*** (2555.34)	17724.39*** (2314.16)	17983.24*** (2323.58)	19889.05*** (2561.86)	19672.79*** (2542.06)
Residual Std. Dev.	29316.68*** (153.33)	29416.26*** (153.94)	29393.40*** (153.78)	29448.81*** (154.37)	29448.93*** (154.35)
N	18323	18303	18314	18241	18246

Notes: Each column shows the expected wage gap for male vs. female, black vs. white, FRPL vs. non-FRPL, ELL vs. non-ELL, and IEP vs. non-IEP student groups. The mean expected wage of pathways follows the definition from Table 1. "Number of pathways offered" is the number of pathways offered by the student's school. The diversity gap is computed by subtracting the diversity indices of the reference group from that of the focus group. Standard Deviations of the Diversity gap for each group are: Male-Female (0.18), Black-White (0.12), FRPL (0.08), IEP (0.12), ELL (0.14). (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$