THREE ESSAYS ON EDUCATION POLICY

Elc Alleb Christian Estrera

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Public Policy.

Chapel Hill
2021

Approved by:

Margaret Burchinal

Thurston Domina

Jane Cooley Fruehwirth

Steven W. Hemelt

Lauren Sartain

©2021
Elc Alleb Christian Estrera
ALL RIGHTS RESERVED

ABSTRACT

Elc Alleb Christian Estrera: Three Essays on Education Policy
(Under the direction of Steven W. Hemelt)

I combine novel data sources with quasi-experimental approaches and detailed knowledge of K-12 policy contexts to explore the causal effects of a range of policies on student cognitive and noncognitive outcomes and teacher labor supply. In Chapter 1, I use event-study and difference-in-differences approaches to estimate the effects of exposure to school active-shooter drills on measures of attendance and achievement. Attendance rates in academic quarters with shooter drills are 0.14-0.16 percentage points (pp) lower than in quarters without such drills. Proficiency rates among 3^{rd}-5^{th} graders exposed to shooter drills before reading and math tests are 0.40 pp lower than that of their counterparts exposed to pre-test fire drills; among 6^{th}-8^{th} graders, the drop is 0.50 pp in math and science. The negative effects of exposure to such drills appear to operate through distress or trauma and not merely disruptions to instructional time.

In Chapter 2, I estimate the causal effect of older-age kindergarten enrollment on K-3 outcomes measured at high-frequency (i.e., 3 or 4 times per school year). My estimation strategy leverages a regression discontinuity approach and controls for measures of pre-kindergarten human capital. During kindergarten and first grade, teachers perceive older and younger kindergarten entrants similarly in their cooperativeness but older entrants as more skilled at working independently. Older kindergarteners enter school with further developed foundational

literacy skills; however, by the end of both kindergarten and first grade, older entrants are similar to younger entrants in these skills. Finally, older kindergarten entrants are as likely as their younger counterparts to be identified with a disability throughout grades K-3.

In Chapter 3, I explore teacher responsiveness to school-level sociodemographic changes resulting from a socioeconomic desegregation plan implemented in Wake County, North Carolina, during the 2000s. In years when schools undergo sociodemographic changes induced by this plan, teachers do not resign from their schools or apply to transfer to other schools at higher rates compared to years when these schools are unaffected by the plan. Teachers also take fewer sick days in the years when their schools are affected by the plan, and even fewer sick days the year after.

For Nina.
Thank you for your patience and understanding.

ACKNOWLEDGEMENTS

I am deeply indebted to my dissertation committee: Peg Burchinal, Thad Domina, Jane Fruehwirth, Steve Hemelt, and Lauren Sartain. Their support and feedback have been invaluable. I am especially grateful to my graduate school adviser, Steve Hemelt, who offered exceptional mentorship along the way. Without his unwavering commitment to my development as a scholar, I might not have produced these essays.

TABLE OF CONTENTS

LIST OF TABLES x

LIST OF FIGURES xii

LIST OF ABBREVIATIONS xiii

CHAPTER 1: ACTIVE-SHOOTER DRILLS AND SCHOOL OUTCOMES 1

1.1 Overview 1

1.2 Introduction 1

1.3 Background 6

1.3.1 Safety Drills in Arkansas Public Schools 6

1.3.2 Active-Shooter Drill Exposure and School Outcomes: Mechanisms 9

1.4 Data and Measures 12

1.4.1 Data 12

1.4.2 Outcomes 14

1.4.3 Test-Drill Temporal Proximity as a Proxy for Drill Exposure Intensity 15

1.5 Empirical Strategy 16

1.5.1 Active-Shooter Drill Exposure and Attendance Rates 16

1.5.2 Active-Shooter Drill Exposure and ACT Aspire Proficiency Rates 18

1.6 Results 23

1.6.1 Summary Statistics 23

1.6.2 Active-Shooter Drill Exposure and Attendance Rates 25

1.6.3 Active-Shooter Drill Exposure and ACT Aspire Proficiency Rates 27

1.6.4 Robustness Checks 37

1.6.5 Falsification Tests 38

1.7 Discussion 40

1.7.1 Limitations 40

1.7.2 Conclusions 41

REFERENCES 42

CHAPTER 2: KINDERGARTEN ENROLLMENT AGE AND HIGH-FREQUENCY 48

2.1 Overview 48

2.2 Introduction 49

2.3 Data 53

2.4 Empirical Strategy 60

2.5 Results 63

2.5.1 Effects on K-1 Literacy 65

2.5.2 Effects on K-1 Classroom Behaviors 68

2.5.3 Effects on K-3 Disability Status 70

2.6 Discussion and Conclusion 72

REFERENCES 75

CHAPTER 3: TEACHER ABSENCES, TRANSFERS, AND TURNOVER UNDER SCHOOL DESEGREGATION: EVIDENCE FROM A CONTEMPORARY STUDENT REASSIGNMENT PLAN 77

3.1 Overview 77

3.2 Introduction 77

3.3 Literature Review 82

3.4 Background 85

3.4.1 WCPSS and School Integration through Student Assignment 85

3.4.2 The Teacher Transfer Process in WCPSS 87

3.4.3 Teacher Absence Policy 93

3.5 Data 93

3.6 Empirical Strategy 94

3.7 Results 98

3.8 Discussion 104

REFERENCES 106

APPENDIX 1: CHAPTER 1, APPENDIX A 108

APPENDIX 2: CHAPTER 1, APPENDIX B 112

APPENDIX 3: CHAPTER 2, APPENDIX A 118

APPENDIX 4: CHAPTER 3, APPENDIX A 127

APPENDIX 5: CHAPTER 3, APPENDIX B 128

LIST OF TABLES

Table 1.1: Descriptive statistics for all reported active-shooter and fire drills, Arkansas traditional public schools, 2016-2019. 7

Table 1.2: Descriptive statistics for all Arkansas public schools, Attendance Sample, and those in ACT Aspire analytic samples, 2016-2019. 24

Table 1.3: Active-shooter drill exposure and same-quarter quarter attendance rates, Arkanas Public Schools, 2016-2019. 26

Table 1.4: Active-shooter drill exposure and proficiency rates, by grade level groupings and ACT Aspire Test subject, Arkansas Public Schools, 2016-2019. 28

Table 1.5: "Difference-in-differences" alternative specifications for the effect of active-shooter drill exposure on proficiency rates, by grade level groupings and ACT Aspire Test subject, Arkansas Public Schools, 2016-2019. 35

Table 1.6: Falsification test regressing prior year ACT proficiency rate on indicator for timing of active-shooter drill relative to test, Arkansas Public Schools, 2016-2019. 39

Table 2.1: Descriptive statistics by analytic sample and window, Wake County Public School System. 58

Table 3.1: Descriptive statistics for full-time classroom teachers, Wake County Public School System, 2004-10. Maximum number of teachers:12,480. 94

Table 3.2: The relationship between reassignment under the Wake Reassignment Plan and school characteristics, Wake County Public School System, 2004-10. 98

Table 3.3: The effect of the Wake Reassignment Plan on teacher outcomes, 2004-10. 99

Table 3.4: The effect of the Wake Reassignment Plan on teacher outcomes, once-lagged effects, 2004-10. 103

Table A1.1: Example test schedules, Dewitt Elementary School, Arkansas, 2017-2019. 108

Table A1.2: ACT Aspire Tests relative to active-shooter drill, grades 3-8 in the event-study ACT Sample, Arkansas Public Schools, 2016-2019. 110

Table A1.3: Differences in mean school-grade and school characteristics, school-grade-subject-year observations, grades 3-8, Arkansas public schools, 2016-2019. 111

Table A2.1: Robustness checks for attendance analysis estimates (Table 1.2) excluding school-year-quarter records associated with 0-minute active-shooter drills. 112

Table A2.2: Robustness checks for active-shooter drill exposure and ACT Aspire Proficiency Rates (Panel A of Table 1.4). 114

Table A2.3: Robustness checks for estimated effect of active-shooter drill exposure on proficiency rates, using fire drill exposure as a comparison condition (Panel B of Table 1.4). 116

Table A3.1: The effects of older kindergarten enrollment age on DIBELS composite scores, OLS estimates. 123

Table A3.2: The effects of older kindergarten enrollment age on the probability of failing to meet or inconsistently meeting Conduct and Work Habits expectations, regression discontinuity estimates. 124

Table A3.3: The effects of older kindergarten enrollment age on identification with any disability or cognitive disability, regression discontinuity estimates. 125

Table A4.1: The relationship between third-grade reading test scores and DIBELS composite scores. 127

Table A5.1: Node-grade reassignment, counts and schools involved, Wake County Public School System, 2000-10. 129

LIST OF FIGURES

Figure 1.1: Active-Shooter Incidents, School Drills, and Internet Search Interest, 2000-2019. 3

Figure 1.2: Theory of change showing pathways through which school active-shooter drill exposure could affect student outcomes. 10

Figure 1.3: Active-shooter drill exposure and ACT Aspire proficiency rates, Arkansas grades 6-8, 2016-2019 31

Figure 2.1: Regression discontinuity in probability of enrolling in kindergarten at an older age relative to peers in same kindergarten cohort. 63

Figure 2.2: DIBELS composite score by benchmark period, grades K-1. 65

Figure 2.3: Quarterly Conduct and Work Habits teacher ratings, grades K-1 68

Figure 2.4: Probability of any disability identification, by benchmark period, grades K-3. 70

Figure 2.5: Probability of cognitive disability identification, by benchmark period, grades K-3. 71

Figure 3.1: Transfer Request Form, 2005 89

Figure 3.2: Sample Transfer Request Form, 2009. 90

Figure A3.1: Covariate smoothness at kindergarten cut-off date, Literacy Sample. 118

Figure A3.2: Covariate smoothness at kindergarten cut-off date, Behavior Sample 119

Figure A3.3: Covariate smoothness at kindergarten cut-off date, Disability Sample 120

Figure A4.4: Probability of learning disability identification, by benchmark period, grades K-3. 121

Figure A5.5: Probability of speech impairment identification, by benchmark period, grades K-3. 122

LIST OF ABBREVIATIONS

ADE	Arkansas Department of Education
AFT	American Federation of Teachers
CCD	Common Core of Data
CMS	Charlotte-Mecklenburg Public Schools
DIBELS	Dynamic Indicators of Basic Early Literacy, Next Edition
ECLS-K	Early Childhood Longitudinal Study, Kindergarten Class of 1998-99
ELL	English Language Learner
FRL	Free- or Reduced-Price Lunch
K-3	Kindergarten – 3rd Grade
KIA	Kindergarten Initial Assessment
LEP	Limited English Proficient
NASP	National Association of School Psychologists
NBCT	National Board Certified Teacher
NBPTS	National Board for Professional Teaching Standards
NCES	National Center for Education Statistics
NEA	National Education Association
OSA	Office of Student Assignment
RD	Regression Discontinuity
SLDS	Statewide Longitudinal Data Systems
SPED	Special Education
TRF	Transfer Request Form
WCPSS	Wake County Public School System

CHAPTER 1: ACTIVE-SHOOTER DRILLS AND SCHOOL OUTCOMES[1]

1.1 Overview

Nearly all K-12 public school students participate in active-shooter drills to receive training on responding to on-campus threats. In recent years, K-12 educators, parents of young children, and the press have raised concerns about potential detrimental effects of these drills on student well-being. Yet limited empirical evidence is available to guide policy debates on school outcomes. I use hand-collected data on the timing of standardized tests to investigate the effects of drill exposure on academic outcomes for students in Arkansas. Using event-study and difference-in-differences approaches, I find that active-shooter drills have modest negative effects on measures of attendance and achievement. Analyses that account for interruptions to student learning associated with any safety drill (e.g., fire drills) suggest that the negative effects of active-shooter drills on achievement operate through channels outside of mere reductions in instructional time. The results have implications for policymakers who intend to keep students safe in life-threatening situations that have become more common—both in general and in schools—over the past two decades.

1.2 Introduction

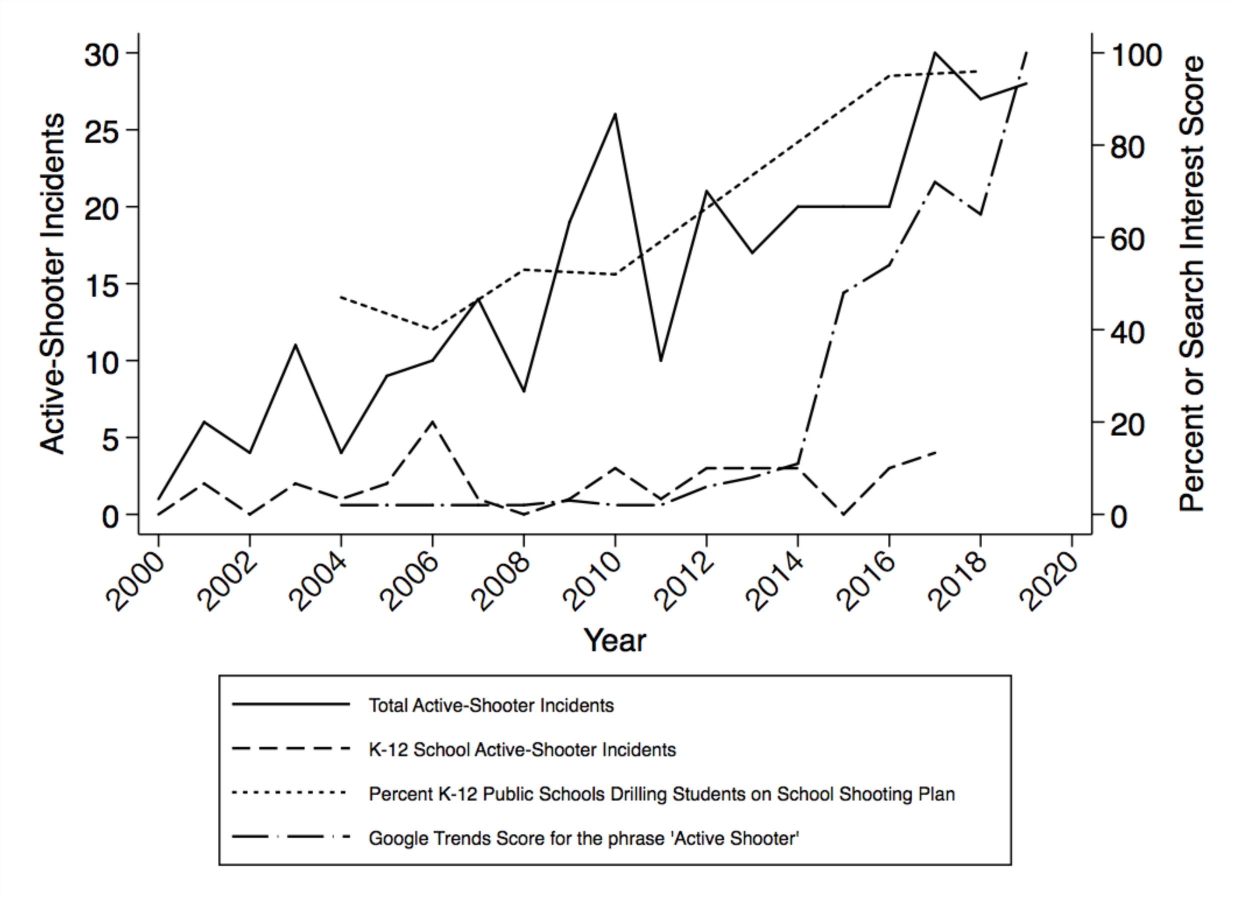
Local and state education agencies institute policies and practices to ensure that students receive training on how to respond to potentially life-threatening incidents. One such incident

[1] This study was reviewed by the University of North Carolina at Chapel Hill Office of Human Research Ethics and on 7/31/2019 determined not to constitute human subjects research as defined under federal regulations [45 CFR 46.102 (e or l) and 21 CFR 56.102(c)(e)(l)] and therefore does not require IRB approval. Study #19-1772.

involves an active shooter attempting to harm people indiscriminately in a confined and populated area. To train students to respond to an active-shooter incident, a school will hold exercises during which staff and students rehearse a set response. The response might require students and staff to lock themselves in classrooms, remain quiet, and wait until the threat is addressed by law enforcement officers (Diliberti et al. 2019; Jonson 2017; Jonson, Moon, and Gialopsos 2020; Jonson, Moon, and Hendry 2018; Musu-Gillette et al. 2018; Schildkraut, Nickerson, and Ristoff 2020). These exercises are known as *lockdown drills.*

More intense versions of these drills are known as *active-shooter drills.* Law enforcement officers might act as "masked gunmen" (Kamenetz 2020) and fire empty shotgun rounds during the drill (Richter 2019), presumably to closely simulate a true shooter threat. In response, students and staff will execute "options-based" steps consistent with the "Run. Hide. Fight." model endorsed by federal agencies (e.g., Department of Homeland Security, Federal Emergency Management Agency) and security training companies (e.g., ALICE Training Institute). Students and staff might flee the scene or barricade themselves in a classroom by stacking chairs and desks "like a fort" (Christakis 2020). They might also include distract and actively resist by throwing objects, swarming, or charging the simulated intruders (Jonson 2017; Jonson, Moon, and Gialopsos 2020; Jonson, Moon, and Hendry 2020; Musu-Gillette et al. 2018; Schildkraut, Nickerson, and Ristoff 2020). I offer the first empirical evidence that active-shooter drills have unintended, negative consequences for school accountability outcomes.

Figure 1.1: Active-Shooter Incidents, School Drills, and Internet Search Interest, 2000-2019.



Notes: Active-shooter incident counts were published by the United States Federal Bureau of Investigation. The percent of K-12 public schools drill students on school active shooter plans come from tables in *Crime, Violence, Discipline, and Safety in US Public Schools*, published by the National Center for Education Statistics, which calculated these estimates based on the *School Survey on Crime and Safety* conducted during the following school years: 2003–04, 2005–06, 2007–08, 2009–10, 2015–16, and 2017–18. The Google Trends Interest Score is the maximum monthly score in a given calendar year for the search phrase “active shooter.”

Figure 1.1 motivates the policy relevance of this paper. As active-shooter incidents have increased in frequency since the year 2000, so too has the number of schools reporting maintaining a written plan for responding to an on-campus shooting or active shooter and drilling students on the plan. Yet school active-shooter incidents make up a small share of all active-shooter incidents over this time period. Concurrently, interest among Internet users for search results containing the phrase "active shooter" has also risen over this time period. Such interest is indicated by each year's maximum monthly Google Trends Search Interest Score, a measure of Internet search popularity, for the phrase "active shooter."

In recent years, educators have raised concerns about training students on how to respond to active shooters. Out of concern that lockdown and active-shooter drills could traumatize students participating in such drills, the National Association of School Psychologists (NASP) recommends that staff should be trained to recognize signs of trauma, student participation should not be mandatory, and student involvement should always require parental consent (2018). The American Federation of Teachers (AFT) and the National Education Association (NEA), two of the largest teachers' unions in the United States, "do not recommend training for students" (Everytown 2020), although the protection of "student well-being" should be a priority if student involvement must occur. The AFT and the NEA also advise against individuals simulating true shooters in drills involving students.

To my knowledge, no study has assessed the relationship between active-shooter drill exposure and school outcomes. To make inroads into this topic, I address two research questions. First, is attendance lower during academic quarters when schools conduct active-shooter drills, relative to quarters when no such drills occur? Second, is exposure to an active-shooter drill during the weeks leading up to an end-of-year test associated with smaller shares of students scoring at least proficiently on the test?

The policy context of the current study is Arkansas. In 2015, the Arkansas General Assembly amended the Safe Schools Initiative Act (AR Code § 06-15-1303). The amendment required every school district in the state to develop a "school safety plan" to prevent and respond to acts of violence and natural disasters. Specifically, the Act requires school districts, beginning in the 2015-16 school year, to provide "annual training for all of its employees and students" in the form of tornado, fire, and active-shooter drills.

I preview the two main results. First, I find that the attendance rate for academic quarters when active-shooter drills occur is less than 1 percent of a percentage point (pp) lower than in quarters without these drills. Second, I document lower proficiency rates on the state's end-of-year achievement tests, the ACT Aspire Summative Assessment (or the ACT Aspire Tests), when these tests are taken after, but in close temporal proximity to, an active-shooter drill, relative to when these tests are taken before the drill. Specifically, I find that the share of middle school students that score at least proficiently on the ACT Aspire math and science tests is 3.75-4 pp lower, respectively, if these tests are taken during the 7 calendar days following an active-shooter drill, relative to these tests taken during the 7 days preceding such a drill. There is no such effect if math and science tests occur 8 or more days after the drill, which implies a rather short-lived effect of these drills on test performance. I mitigate the possibility that losses in instructional time and student learning are the mechanism through which active-shooter drill participation lowers proficiency rates. I recover modest but significant negative effects of drill exposure on reading and math proficiency rates for elementary schools as well as on math and science proficiency rates for middle schools. I contend that the unique aspects of these drills have distinct negative effects on test scores over and above mere reductions in instructional time and student learning.

1.3 Background

1.3.1 Safety Drills in Arkansas Public Schools

Students in Arkansas public schools participate in three types of safety drill throughout the year: tornado, fire, and active-shooter. I focus on active-shooter and fire drills, because at least one of these drills occurs at least once per month of the school year, including April and May when schools administer end-of-year tests. The state requires that every school conducts one active-shooter drill per school year. I find that 92 percent of schools report conducting one active-shooter drill per year during the 2016-19 school years inclusive;[2] 5 percent of schools, 2 such drills per year; and the remaining schools report conducting 3 or more such drill per year.

Table 1.1 summarizes the timing and the duration of fire and active-shooter drills over the years 2016-19. Active-shooter and fire drills are distributed similarly across academic quarters. For each drill type the majority occur before the first day of the state and school testing windows or after the last day. Historically, the state's ACT Aspire testing window spans the first Monday of April through the second Friday of May. Both types of safety drill are distributed similarly across days of the week. Active-shooter drills more often than fire drills occur in the morning.

[2] I use the calendar year of the spring term to refer to the school year, e.g., 2017 refers to the 2016-17 school year.

Table 1.1: Descriptive statistics for all reported active-shooter and fire drills, Arkansas traditional public schools, 2016-2019.

	Active-Shooter		Fire	
	Proportion	Obs	Proportion	Obs
Before First Day of School	0.05	2522	0.00	21049
After Last Day of School	0.01	2522	0.00	21049
Quarter				
1	0.22	2522	0.23	21049
2	0.19	2522	0.27	21049
3	0.24	2522	0.24	21049
4	0.28	2522	0.26	21049
Between End of One Quarter / Start of Next Quarter	0.03	2522	0.03	21049
Within 2 Weeks of Academic Quarter Start	0.19	2522	0.22	21049
Within 2 Weeks of Academic Quarter End	0.24	2522	0.25	21049
Relative to State ACT Aspire Window				
Before First Day	0.78	2522	0.82	21049
After Last Day	0.08	2522	0.05	21049
During	0.14	2522	0.13	21049
Relative to School ACT Aspire Window				
Before First Day	0.82	2522	0.86	21049
After Last Day	0.16	2522	0.12	21049
During	0.02	2522	0.02	21049
Day				
Sun	0.01	2522	0.00	21049
Mon	0.17	2522	0.17	21049
Tue	0.20	2522	0.21	21049
Wed	0.22	2522	0.20	21049
Thu	0.21	2522	0.21	21049
Fri	0.19	2522	0.20	21049
Sat	0.00	2522	0.00	21049
Morning Drill	0.67	2522	0.57	21049
Drill Duration (Minutes)^				
0	0.16	361	0.00	21049
(0,1]	0.12	361	0.14	21049
(1,2]	0.21	361	0.51	21049
(2,5]	0.25	361	0.28	21049
(5,10]	0.10	361	0.03	21049

(10,15]	0.04	361	0.01	21049
(15,20]	0.03	361	0.00	21049
(20,25]	0.01	361	0.00	21049
(25,30]	0.01	361	0.00	21049
>30	0.06	361	0.02	21049

Notes: The unit of observation is the drill-year. A drill is classified as "between quarters" if it occurs: (a) after the 1st quarter end date and before the 2nd quarter start date; (b) after the 2nd quarter end date and before the 3rd quarter start date; or (c) after the 3rd quarter end date and before the 4th quarter start date. ^Active-shooter drill duration reported only for 2016.
Source: Arkansas Department of Eduation Data Center.

The reported durations of active-shooter drills and fire drills are suggestive of the magnitude of lost instructional time and reduced student learning associated with exposure to these drills. Importantly, the reported durations of active-shooter drills in Arkansas are consistent with those reported in the only other study documenting lockdown and active-shooter drill durations. Active-shooter drills in classrooms have a median duration of 15 seconds (Jonson, Moon, Leo 2020). When active-shooter drills are conducted in large open area (e.g., a cafeteria), the median drill duration rises slightly to 150 seconds, or about 2-3 minutes. This examination of drill duration suggests that, on average, students participating in active-shooter drills lose more instructional time—but not considerably more—than students participating in fire drills.

While the ADE Data Center contains ample information on the timing of safety drills, the data do not describe how students and staff respond to these drills (e.g., barricade doors, hide, resist), whether law enforcement officers act as intruders, or the extent to which these officers take actions to simulate true shooter threats (e.g., fire empty shotgun rounds). Some evidence suggests that the implementation of school active-shooter drills is uniform across the United States, including across Arkansas. Some school districts outsource the process of conducting active-shooter drills to third-party security training companies. The ALICE Training Institute is one prominent company. The acronym ALICE denotes a modified form of the steps “Run. Hide.

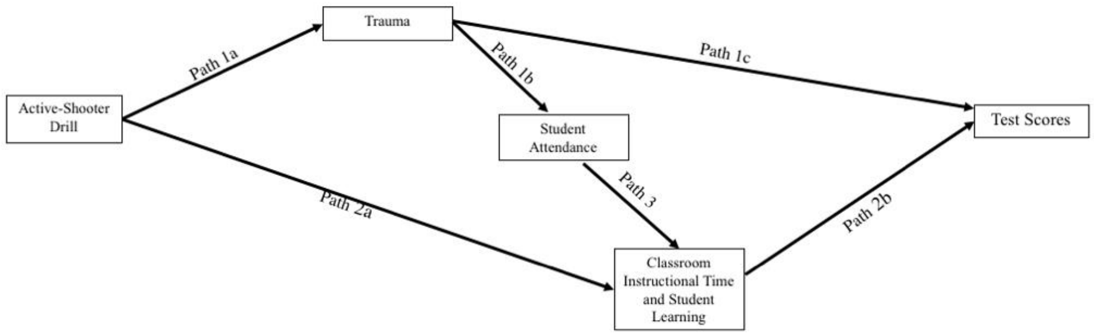
Fight.": Alert, Lockdown, Inform, Counter, and Evacuate (Alice Training Institute 2020). The ALICE Training Institute's customer base of school districts has drawn the attention of the press (e.g., O'Regan 2019). At the time of this writing, the company reported serving 5,548 school districts (Alice Training Institute 2020), which is about 30 percent of the 19,400 school districts across. In Arkansas specifically, the state's Bureau of Legislative Research (2014) reported that by 2014 at least 50 school districts received some form of ALICE Training. Figure 1.1 suggests that that estimate has probably increased.

Anecdotes in the press and descriptions in empirical studies constitute what is known about the experiential features of school active-shooter drills generally. Nevertheless, my study rests on two rather weak assumptions about these drills. The first is that public schools in Arkansas conduct active-shooter drills. Indeed, the ADE Data Center labels these drills as such. The second is that public school students in Arkansas participate in active-shooter drills. The distribution of active-shooter drill dates across weekdays mirrors that of fire drill dates. Likely all students participate in fire drills.

1.3.2 Active-Shooter Drill Exposure and School Outcomes: Mechanisms

Figure 1.2 presents a theory of change for how active-shooter drill exposure could influence school outcomes. There are two mechanisms, which I describe in turn.

Figure 1.2: Theory of change showing pathways through which school active-shooter drill exposure could affect student outcomes.



Trauma is the first mechanism through which active-shooter drill exposure could lower school outcomes (Path 1a). This is the mechanism that appears to have inspired the policy recommendations put forth by the AFT and the NEA, as well as cautionary coverage in the popular press about active-shooter drills. Empirical studies relevant to this mechanism examine the relationship between students' perceptions of safety at school exposure to lockdown drills, active-shooter drills, and variants thereof. Drill variants include discussion-based protocols (Jonson, Moon, and Gialopsos 2020), retrospective surveys on prior drill participation and perceived safety (Huskey and Connell 2020), and training videos (Peterson et al. 2015). Participants in these studies varied in grade-level, ranging from elementary- and secondary- (Jonson, Moon, and Gialopsos 2020; Schildkraut, Nickerson, and Ristoff 2020; Zhe and Nickerson 2007) to university-age students (Huskey and Connell 2020; Peterson et al. 2015). The results in these studies are mixed. On the one hand, students' perceived safety following exposure to drills or drill variants declines (Huskey and Connell 2020; Peterson et al. 2015;

Schildkraut, Nickerson, and Ristoff 2020). On the other hand, students' self-reported perceived safety at school post-exposure is unaffected (Jonson, Moon, and Gialopsos 2020; Zhe and Nickerson 2007).

Exposure to putatively traumatic events can lower attendance (Path 1b). Following a fatal sniper attack on an elementary school playground, student absences were significantly higher one week after the shooting relative to the week before, and remained at these levels for one month after (Pynoos et al. 1987). Daily attendance rates fell to as low as 10 percent at elementary schools close to one of the Beltway Sniper shooting sites (Gershenson and Tekin 2018). Shootings at public schools lower enrollment at these schools and raise private school enrollment (Abouk and Adams 2013). These forms of absenteeism and avoidance are consistent with the *Diagnostic and Statistical Manual of Mental Disorders* (5^{th} ed., DSM-V) listing avoidance behaviors as one of four diagnostic criteria clusters under Posttraumatic Stress Disorder.

Exposure to putatively traumatic events can also lower test scores (Path 1c). The basic result in studies by Patrick Sharkey and coauthors is that increased temporal and physical proximity to neighborhood violence, especially homicides, has large negative effects on achievement (e.g., reading and vocabulary tests) as well as noncognitive outcomes (e.g., impulse control, attention) (Sharkey 2010; Sharkey et al. 2012, 2014). Gershenson and Tekin (2018) found that among third and fifth graders attending public schools close to the Beltway Sniper Attacks, 2-5 percent fewer of these students scored proficient on the Virginia Standards of Learning end-of-year tests relative to their peers who attended schools farther away from these sites. Beland and Kim (2016) reported that students attending high schools in years when homicidal shootings occurred on campus scored 3.9 and 4.9 percentage points lower in English and math, respectively, relative to years with no school shootings.

Reductions in classroom instructional time and student learning (Path 2a) form the second mechanism through which active-shooter drill exposure lowers test scores (Path 2b). When severe weather closes schools, students score lower on end-of-year tests than they would have had schools been open (Goodman 2014; Marcotte and Hemelt 2008). Conversely, when students attend school for more days than they would have otherwise, they score higher on tests (Aucejo and Romano 2016; Liu, Lee, and Gershenson 2019). Active-shooter drill exposure on the days preceding these tests probably reduces the time that students and teachers might otherwise spend "cramming" (Donovan, Figlio, and Rush 2006). How safety drills lower instructional time also fits into a broader category of classroom interruptions (e.g., unanticipated intercom announcements) (Kraft and Monti-Nussbaum 2020). Notably, these interruptions include the residual effects of drill participation. Students spend time and cognitive energy shifting their attention post-drill to the resumption of pre-drill tasks (Altman and Trafton 2004; Altman, Trafton, and Hambrick 2014). As such, they incur setbacks in knowledge acquisition and information recall (Foerde, Knowlton, and Poldrack 2006; Gillie and Broadbent 1989).

There is also an indirect path through which active-shooter drill exposure lowers instructional time. Although my results suggest the relevance of this path is quite weak, active-shooter drills might lead some students to stay home from school (Path 3), which would reduce classroom instructional time for these students and potentially lower their test scores.

1.4 Data and Measures

1.4.1 Data

I began by compiling publicly available accountability data pertaining to traditional K-12 public schools in Arkansas. The sources for these data are the ADE Data Center and the National Center for Education Statistics (NCES). These sources report information at an aggregate level (i.e., the school or school-grade unit) and at a particular frequency (i.e., quarterly or annually).

The ADE datasets are the primary source of school and school-grade information in my study, including sociodemographics (e.g., percent racial/ethnic minority), safety drill dates and durations, quarterly attendance rates, and ACT Aspire proficiency rates. I merged into the ADE data school-level, urban-centric locale codes from the NCES Common Core of Data (CCD), which classify the location of each school as city, suburb, town, or rural.

To investigate the relationship between active-shooter drill exposure and achievement, I augmented the ADE and NCES data by merging in ACT Aspire test schedules for 2016-19. A research assistant who was a legal resident of Arkansas submitted a public records request to the ADE for test schedules. The recovered test schedules varied in layout and file format. I recruited a team of research assistants to key test dates into spreadsheets, which resulted in a dataset whose unit of observation is the school-grade-subject-year (e.g., 3rd graders at elementary school X testing in math in 2017). ACT Aspire test schedules documented either: (a) the test date for every observation; or (b) a testing window, meaning the two dates between which a school-grade-year unit (e.g., 3rd graders at elementary school X in 2017) completed one or more ACT Aspire subject tests. I corrected errors in the keyed-in test dates (e.g., wrong school year), which were exceedingly minimal, likely due to the simplicity of the coding task. Combined with active-shooter drill dates in the ADE data, every test schedule allowed me to determine for every school-grade-year unit the timing of an active-shooter drill relative to every ACT Aspire test, with either day- or week-level precision but not both.

Table A1.1 (Appendix 1) presents an example set of recovered test schedules for grades 3-5 at Dewitt Elementary School in Arkansas's Dewitt School District. Columns 1-4 show the unique NCES School Identifier, grade level, ACT Aspire Test subject, and school year, respectively. For this elementary school, I recovered test schedules for the years 2017-19 inclusive; I recovered no test schedules for 2016. Columns 5 and 6 show the start and end date for the test. If a test schedule noted the exact test date rather than a test window, Columns 5 and 6 equal one other and Column 7 is equal to 1.

On the whole, I recovered test schedules for 68 percent of all grade 3-10 school-grade-year observations associated with traditional public schools and active-shooter drill dates for the school years 2016-19. This amounts to 83 percent of all unique traditional public schools for which the ADE reported at least one active-shooter drill during this time period. In the Results section below, I offer descriptive evidence that the school-year observations for which I recovered at least one test schedule are similar on observable characteristics to all school-year observations over this time period.

1.4.2 Outcomes

I estimate the effect of active-shooter drill exposure on up to five outcome variables: (1) the attendance rate, and (2) the proficiency rates on the ACT Aspire Summative Assessment in English, reading, math, and science. The attendance rate is a school-level measure equal to the total number of days that students are in school (average daily attendance) divided by the total number of days that students could have been in school (average daily membership). I compute the quarterly attendance rate for every school over the school years 2016-19.

Beginning in 2016, Arkansas administered the ACT Aspire Tests to all students in grades 3-10. The ACT Aspire Tests are a psychometrically validated, vertically scaled battery of achievement tests designed to measure growth for students in grades 3-10. The same developers of the widely administered college admissions test known as the ACT also developed the ACT Aspire Tests. As such, the ACT Aspire Tests assess college readiness according to the ACT College Readiness Benchmarks (Allen and Radzunel 2017; ACT 2018, 2019). Students scoring "at or above [the College Readiness Benchmarks] are on target for college readiness when they take the ACT test in grade 11" (ACT 2018, p. 3), and the ACT College Readiness Benchmarks predict performance in college courses (ACT 2018, 2019). I note that while the battery of tests includes a Writing subject test, I exclude this test from my analyses because the scaling for the Writing test scores is based on a rubric, meaning there is no vertical scale interpretation (ACT Research Services 2019). For each ACT Aspire subject test, the ADE reports at the level of the school-grade-year the percentage of students scoring at or above the benchmark, or at least proficiently, for that grade-subject combination.

I restrict the achievement analysis to grades 3-8 for two reasons. First, there are more elementary and middle schools than high schools in the state, allowing me to form the largest samples using these grade levels. Second, the well-being of older students is less likely affected by presumably traumatic events, for example, school shootings (Beland and Kim, 2016), compared to younger students (Pynoos et al. 1987).

1.4.3 Test-Drill Temporal Proximity as a Proxy for Drill Exposure Intensity

The ACT Aspire test schedules allow me to derive a key measure in my study: *test-drill temporal proximity*, or how close in time—in weeks—is the test date to the date of the active-shooter drill. A characteristic of all safety drills, test-drill temporal proximity takes on positive and negative integer values. Values increasing in magnitude in either direction imply testing further away in time from the drill. Negative and positive integers approaching 0 imply testing close in time to the drill. The value 0 for test-drill temporal proximity denotes testing the same day as the drill; I exclude these observations from my analysis because I cannot tell when the test occurred relative to the drill on the drill day. The value 1 denotes testing the day after the drill day; the value -1, testing the day before the drill day.

Test-drill temporal proximity operationalizes a key measure in the current study, that is, *drill exposure intensity*, or how much drill participation plus the drill's residual effects are presumed to affect ACT Aspire subject test proficiency rate. There is a precedent for this kind of exposure measure. It is similar in spirit to the way that studies have operationalized exposure to: neighborhood violence (Sharkey 2010; Sharkey et al. 2012, 2014), threats to students' physical safety (Gershenson and Tekin 2018), and on-campus school shootings (Abouk and Adams 2013; Beland and Kim 2016; Rossin-Slater et al. 2020). These studies measure an individual's exposure to the life-threatening event in terms of the individual's temporal or physical proximity to the event. I use only temporal proximity because physical proximity in the current study is effectively zero; presumably students are on-campus during the safety drill. If safety drill exposure has any effect on the ACT Aspire proficiency rate, then the stronger the effect of the drill on the proficiency rate the closer is the safety drill to the ACT Aspire Test.

1.5 Empirical Strategy

1.5.1 Active-Shooter Drill Exposure and Attendance Rates

If active-shooter drills lower student well-being, then I expect that the attendance rate for quarters when active-shooter drills occur will be lower than for quarters when no such drills occur, or when only fire and tornado drills occur. I estimate the following equation:

$$(1)\ AttendanceRate_{stq} = \alpha_s + \beta ShooterDrill_{stq}$$
$$+X_{st} + \sum_t \phi_t + \sum_q \chi_q + \sum_t \sum_q \mu_{tq} + \sum_s \sigma_s + \sum_s \sum_q \psi_{sq} + \varepsilon_{stq},$$

which is indexed by school s, year t, and quarter q. *AttendanceRate* denotes the quarterly attendance rate for a specific school-year-quarter. $ShooterDrill$ denotes an indicator variable equal to 1 if one or more active-shooter drills occurred during a specific school-year-quarter; 0 if no active-shooter drill occurred, or equivalently only fire or tornado drills occurred. X denotes time-varying, school-level characteristics, including: the percent of students economically disadvantaged (i.e., eligible for free or reduced-price lunch), the percent of Black students, the percent of Hispanic students, and the student-teacher ratio. The terms ϕ, χ, μ, σ, and ψ denote year, quarter, year-by-quarter, school, and school-by-quarter fixed effects, respectively. ε_{stq} denotes the stochastic error term.

The parameter of interest, β, represents the relationship between active-shooter drill exposure and the quarterly attendance rate. β denotes how much the attendance rate for the same school-quarter differs between years. The identifying variation to estimate β comes from changes in active-shooter drill timing within-school and between quarters. That is because active-shooter drills occur in the school during a particular quarter in some years, and in other years only fire and tornado drills occur in this same school during that same quarter.

Two limitations stem from the fact that the ADE reports the attendance rate at a quarterly frequency and not a daily frequency. First, I cannot distinguish between fluctuations in the attendance rate due to students who are not in school on days preceding or following the drill. Whereas some students who might get wind of the date of an active-shooter drill could stay home to avoid it, other students might stay home the day after to avoid reminders of the discomfort they might have felt during the drill. Second, I cannot address the fact that active-shooter drill exposure during the final weeks of a quarter almost certainly influences the subsequent quarter's attendance rate. By example, active-shooter drill exposure on the last day of quarters 1, 2, or 3 is likely to affect the attendance rate of quarter 2, 3, or 4, respectively, if students stay home following drill exposure.

While addressing the first limitation would require day-level attendance rates, the data permit me to address the second limitation using a two-pronged approach in which I re-estimate Equation (1). First, I exclude from the analytic sample school-year-quarter observations in quarters 2-4 if I observed active-shooter drills conducted during the last one or two weeks of the preceding quarters (i.e., quarters 1-3). The attendance rate for these observations could be affected, at least in part, by end-of-quarter active-shooter drills occurring during the preceding quarters. Second, I exclude school-year-quarter observations in quarters 1-3 if I observed active-shooter drills during the last one or two weeks of these same quarters. Active-shooter drills occurring during these weeks of the quarter might not affect the same-quarter attendance rate. This approach reduces contamination in the estimated effect of active-shooter drill exposure during a particular quarter on that same quarter's attendance rate.

1.5.2 Active-Shooter Drill Exposure and ACT Aspire Proficiency Rates

To estimate the effect of active-shooter drill exposure on ACT Aspire proficiency rates, I begin by estimating the following event study model:

$$(2)\, y_{sgt} = \alpha + \sum_s \sum_g \sum_t TestWeek_{sgt} + \tau X_{sgt} + \gamma Q_{st} + \sum_s \sum_g \sigma_{sg} + \sum_t \tau_t + \varepsilon_{sgt},$$

which is indexed by school *s*, grade *g*, and year *t*. *y* denotes the school-, grade-, and year-specific proficiency rate for the English, reading, math, or science ACT Aspire test. I estimate Equation (2) for each subject test separately. *TestWeek* operationalizes drill exposure intensity in the form of mutually exclusive indicator variables, each of which equals 1 for the week (i.e., span of 7 days) of the test relative to the drill, 0 otherwise. The *TestWeek* reference indicator denotes testing during the 7 days leading up to the drill, meaning the reported estimates for all other *TestWeek* indicators are relative to testing during the 7 days leading up to the drill. X denotes a vector of time-varying school-grade covariates, namely the percentages of Black and Hispanic students. Also included in X is a data quality indicator variable for test-date precision. This variable equals 1 for a school-grade-subject-year (e.g., 3rd graders in school X taking the math test in 2016) if the recovered test schedules indicate a specific test date for that school-grade-subject-year; 0, if the test schedules referred to a testing window, that is, to two dates between which the test might have been administered. Q denotes time-varying school covariates, namely the student-teacher ratio and the percentage of economically disadvantaged students, or those eligible for free- or reduced-price lunch. σ and τ denote school-by-grade and year fixed effects, respectively. ε_{sgt} is the stochastic error term.

There is no straightforward way to incorporate the small share of school-year records—8 percent of all such records—where each is associated with more than one active-shooter drill per year. For each such school, it is unclear which active-shooter drill affects proficiency rates. Rather than drop these school-year records from the event study analysis, I treat the first reported active-shooter drill for each such record as its only active-shooter drill. I do so because this particular drill is the first in the year to which students are exposed and as such is probably unannounced. While this approach requires me to discard data, dropping these school-year records raises the rigor of the event study analysis by biasing downward the estimated effect.

To build on the event study model, I estimate a separate model that allows me to compare pre-test active-shooter drill exposure to pre-test fire drill exposure as the counterfactual condition. This approach allows me to test the Classroom Instructional Time Mechanism described in the theory of change above. In some years, a school-grade-subject-year unit tests after an active-shooter drill and in other years after a fire drill. I use between-year variation in pre-test drill-type exposure to recover an estimate of active-shooter drill exposure on proficiency rates that nets out the disruptive qualities common to both active-shooter and fire drills. To be clear, I assume that fire drills are not traumatic in ways that active-shooter drills might be. If fire drills lower proficiency rates, then presumably they do so only through the reduction in instructional time and associated disruptions to student learning.

For this particular analysis, I make the following simplifying assumption: that the safety drill preceding closest in time to an ACT Aspire test is the drill that affects that test's proficiency rate. This assumption is consistent with the drill exposure intensity measure I described above. The idea is that a drill preceding closest in time to the test subtracts away instructional time that is probably more important to student learning than the same amount of instructional time positioned early in the school year. Under this assumption, I classify each school-grade-subject-year observation into one of two groups. In the first, every observation is exposed to an active-shooter drill before the test, and no fire drill occurs after the active-shooter drill and before the test. I label this the *active-shooter drill exposure group*. In the second group, every observation is exposed to a fire drill before the test, and no active-shooter drill occurs after the fire drill and before the test. I label this the *fire drill exposure group*.

I estimate the following equation:

$$(3)\ y_{sgjt} = \beta_0 + \beta_1 SDrillGroup_{sgt} + \beta_2 DrillExposureIntensity_{sgt} + \beta_3\left(SDrillGroup_{sgt} \text{ x } DrillExposureIntensity_{sgt}\right) + \eta X_{st} + \kappa G_{sgt} + \sum_s \sum_g \sigma_{sg} + \sum_t \tau_t + \varepsilon_{sgt}.$$

y denotes the proficiency rate for school *s*, grade *g*, subject *j*, and year *t*. I estimate this equation and report estimates for each ACT Aspire subject test separately. *SDrillGroup* denotes an indicator variable equal to 1 if the school-grade-subject-year unit is in the active-shooter drill exposure group; 0, the fire drill exposure group. *DrillExposureIntensity* denotes a continuous variable taking on the values 1-32 inclusive, where 1 denotes a drill occurring anytime during the 32^{nd} calendar week (7 days) prior to the test, and 32 denotes a drill occurring anytime during the 1^{st} calendar week preceding the test. All other terms are the same as those in Equation (2).

Equation (3) is structurally similar to a difference-in-differences (DD) estimator, where the active-shooter drill exposure group is the treatment group and the fire drill exposure group is the comparison group. Therefore, the parameter of interest, β_3, denotes the change in the proficiency rate associated with a 1-week increase in drill exposure intensity among school-grade-subject-year units exposed to an active-shooter drill before the test relative to school-grade-subject units exposed to a fire drill before the test. In other words, β_3 is the unique effect of active-shooter drill exposure on proficiency rates that exists over and above drill-induced instructional time loss and reductions in student learning, which are accounted for in fire drill exposure. If the magnitude of the estimate for β_3 is 0, then lost instructional time and disruptions to student learning explain the entirety of the negative effect of pre-test active-shooter drill exposure on proficiency rates.

I extend the DD-style analysis in two ways. First, I relax the assumption that the effect of active-shooter drill exposure on proficiency rates fades out in a linear pattern. I re-estimate Equation (3) and add to the right hand-side $DrillExposureIntensity^2$ and the interaction term $SDrillGroup \times DrillExposureIntensity^2$. These terms allow me to model the unique effect of active-shooter drill exposure on proficiency rates using a more flexible functional form. Perhaps the fadeout of drill exposure fades out more precipitously than can be modeled linearly.

Second, I interrogate the assumption that the safety drill preceding closest in time to an ACT Aspire test is the drill that affects that test's proficiency rate. While I make this assumption to test the Classroom Instructional Time Mechanism, one objection is that the assumption results in discordant assignments of school-grade-subject-year observations into either the active-shooter or fire drill exposure group. An example illustrates this point. Observation (Obs) 1 is exposed to a fire drill 14 days before the test and to an active-shooter drill 7 days before; therefore, Obs 1 is classified into the active-shooter drill exposure group. Obs 2 is exposed to a fire drill 3 days before the test and to an active-shooter drill 4 days before; therefore, Obs 2 is classified into the fire drill exposure group. However, Obs 1 and Obs 2 appear similar to one another. Each is exposed to an active-shooter drill during the 7 days preceding the test. Active-shooter drill exposure intensity—both in instructional time loss and in trauma-inducing potential—is also probably larger for Obs 2 than Obs 1, because of how much closer to the test the Obs 2 active-shooter drill occurs compared to the Obs 1 active-shooter drill. Perhaps Obs 2 should be classified into the active-shooter drill exposure group, not the fire drill exposure group.

I interrogate Equation (3)'s simplifying assumption by determining how much the classification of certain observations into either the fire drill or active-shooter drill exposure group affects the estimated effect of active-shooter drill exposure on proficiency rates. I do so in two ways. First, I drop any observation for which both an active-shooter drill and a fire drill occur before the test and in relatively close proximity to one another. Specifically, I drop any observation for which both an active-shooter drill and a fire drill occur before the test and within 7, 21, or 35 days of one another. As I show in the Results section below, the event-study results imply a negative effect of active-shooter drill exposure on proficiency rates for school-grade units that test within 7 days after such a drill. Therefore, if both a fire drill and an active-shooter drill precede a test and are 7 days or less apart from one another, then the active-shooter drill's effects on proficiency rates—by inducing trauma or reducing instructional time and student learning—may very well contaminate the effects of the fire drill on proficiency rates. By dropping observations for which active-shooter and fire drills occur within 7 days of one another, I reduce such contamination. As I increase this timeframe to 21 and 35 days, I reduce contamination even more for a reasonable loss in statistical power. To be clear, I drop even the observations for which both the fire and active-shooter drill occur before the 7 days preceding the test. Doing so allows me to restrict the analytic sample to instances of active-shooter drills whose unique effects on proficiency rates I am probably able to isolate.

Second, instead of dropping certain school-grade-subject-year observations, I classify into the active-shooter drill exposure group any such observation if it is exposed to an active-shooter drill during the 7, 21, or 35 days preceding the test. To be clear, I reclassify these observations even if a fire drill occurs after the active-shooter drill and before the test, which under the initial simplifying assumption would place the observation in the fire drill exposure group. If coefficient estimates deviate in sign or magnitude from those recovered using the initial DD model (i.e., Panel B of Table 1.4), then the exposure group into which I classify these particular observations would seem to matter in critical ways for the estimated effect of active-shooter drill exposure on proficiency rates.

1.6 Results

1.6.1 Summary Statistics

I report summary statistics for traditional public schools (TPS) in Arkansas serving grades 3-10, over the school years 2016-2019. Because the ADE reports active-shooter drill dates for Arkansas TPS, summary statistics for the Attendance Sample are identical to those for Arkansas TPS reporting active-shooter drill dates during this time period. In contrast, the two ACT Samples consist of subsets of Arkansas TPS. I recovered ACT Aspire test schedules for at least one grade-subject unit in each of these schools.

Table 1.2: Descriptive statistics for all Arkansas public schools, Attendance Sample, and those in ACT Aspire analytic samples, 2016-2019.

	(1)	(2)	(3)	(4)	(5)	(6)
			ACT Samples (Grades 3-8)			
	Attendance Sample (Grades 3-10)		For Event-Study Models (i.e., No Fire Drills)		For Models Incorporating Fire Drill Exposure	
	Mean	SD	Mean	SD	Mean	SD
School Percent Black	21.02	27.28	22.39	27.86	22.15	28.04
School Percent Hispanic	12.57	14.54	13.89	15.38	13.16	15.18
School Percent White	61.13	29.54	57.97	29.73	59.19	30.42
School Percent Proficient or Above on ACT Aspire Tests	48.52	14.07	51.25	13.66	50.29	13.90
School Percent Economically Disadvantaged	63.51	19.23	64.88	19.80	64.53	19.75
School Attendance Rate	94.42	2.01	94.55	1.55	94.46	1.95
School Student-Teacher Ratio	12.31	3.90	13.63	3.00	12.97	3.71
Percent Schools in Cities	28.44	-	33.33	-	32.92	-
Percent Schools in Suburbs	13.25	-	14.95	-	13.48	-
Percent Schools in Towns	18.18	-	17.14	-	15.26	-
Percent Schools in Rural Areas	40.13	-	34.58	-	38.34	-
N School-Year	1804		1365		1291	
N School	740		549		699	

Note: The unit of observation is the school-year. Percent proficient or above on ACT Aspire Tests is calculated across all grades in the school, weighted by the number of test-takers in the school-grade-year-subject cell, and across English, reading, math, and science. Economic disadvantage denotes the share of students eligible for free or reduced price lunch. Columns (1) and (2) denote traditional public schools in Arkansas serving grades 3-10, which are included in the Attendance Sample. Columns (3)-(6) denote traditional public schools serving grades 3-8 included in the ACT analytic samples.
Source: Arkansas Department of Eduation Data Center

Table 1.2 shows that the schools in the Attendance Sample and the two ACT Samples are similar to one another on observable characteristics. Some differences are worth noting, namely that schools in the ACT Samples serve slightly smaller shares of White students, slightly higher achieving students (based on the proficiency rate across all tests), and a slightly larger share of economically disadvantaged students.

1.6.2 Active-Shooter Drill Exposure and Attendance Rates

Table 1.3 reports the coefficient estimate on *ShooterDrill* from Equation (1). Academic quarters in which active-shooter drills occur have a 0.16 percentage point (pp) lower attendance rate relative to quarters with no such drills (Column 1). That amounts to an extremely small fraction of the mean attendance rate for quarters with no active-shooter drills. It is also the combined effect of absences on the days leading up to and following active-shooter drills.

Table 1.3: Active-shooter drill exposure and same-quarter quarter attendance rates, Arkansas Public Schools, 2016-2019.

	(1)	(2)	(3)
	All Observations	Exclude End-of-Quarter Shooter Drills	
		Exclude Last 1 Week	Exclude Last 2 Weeks
ShooterDrill=1	-0.160**	-0.147*	-0.141*
	(0.078)	(0.077)	(0.079)
Non-Shooter-Drill Quarter Mean Attendance Rate	94.47	94.48	94.49
R Squared	0.776	0.778	0.778
N School-Year-Quarter	7196	6988	6844
N School-Year	1805	1805	1805
N School	740	740	740

Note: The unit of observation is the school-year-quarter. ShooterDrill denotes the indicator equal to 1 if an active-shooter drill occurred during that quarter. Column (1) include all school-year-quarter observations. Column (2) excludes (a) school-year-quarter observations in quarters 2, 3, and 4 if active-shooter drill exposure occurred during the last week at the end of quarters 1, 2, or 3; and (b) school-year-quarter observations in quarters 1, 2, and 3 if active-shooter drill exposure occurred during the last week at the end of quarters 1, 2, and 3. Column (3) extends the Column (2) exclusion criteria to two weeks. All specifications control for school-level percent economically disadvantaged, percent Black, percent Hispanic, student-teacher ratio, year, quarter, school-by-quarter, and year-by-quarter fixed effects. Estimates are weighted by the number of students in the school-year cell. Heteroskedastic-robust standard errors are in parentheses and clustered on the school. Asterisks denote statistical significance: *p<0.1 **p<.05 ***p<.01.
Source: Arkansas Department of Eduation Data Center

Column 2 excludes school-year-quarter observations: (a) from quarters 1, 2, and 3 if an active-shooter drill occurred during the last week of each quarter; and (b) from quarters 2, 3, and 4 if an active-shooter drill occurred during the last week of each quarter 1, 2, and 3, respectively. Column 3 extends the exclusion criterion in Column 2 from 1 week to 2 weeks. The point estimates in Columns 2 and 3 imply that the attendance rate during quarters when active-shooter drills occur is about 0.14-0.15 pp lower relative to quarters without such drills. Compared to Column 1, Columns 2 and 3 may report more accurate estimates of the effect of active-shooter drill exposure on quarterly attendance rates through the exclusion of active-shooter drills whose effects on attendance rates might only be realized in subsequent quarters. These point estimates are similar in magnitude and the same in direction as the point estimate reported in Column 1. These results imply that active-shooter drill exposure during a particular quarter is associated with a drop in the attendance rate during that quarter.

1.6.3 Active-Shooter Drill Exposure and ACT Aspire Proficiency Rates

Equation (2) is the event study model in which the estimated effect of active-shooter drill exposure on proficiency rates is identified off of variation in test-drill temporal proximity within school-grade units. Table A.2 summarizes the characteristics of school-grade observations by whether they are “switchers” or whether they always test before the drill or always after the drill. Switchers differ from the other two groups in terms of school-grade racial composition, student-teacher ratios, grade levels, and their distribution across urbanicity locales. Switchers are similar to the other two groups in terms of proficiency rates, economic disadvantage, attendance rates, and principal turnover.

Table 1.4: Active-shooter drill exposure and proficiency rates, by grade level groupings and ACT Aspire Test subject, Arkansas Public Schools, 2016-2019.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Grades 3-5				Grades 6-8			
	English	Reading	Math	Science	English	Reading	Math	Science
Panel A. Event Study Estimates								
TestWeek=Week after Drill	-0.61	-2.19	-0.39	-0.14	0.47	-1.33	-4.54**	-3.49*
	(2.02)	(1.79)	(2.17)	(1.96)	(1.75)	(2.33)	(1.96)	(1.82)
Mean Proficiency Rate Test Week before Drill	74.03	40.19	52.26	42.30	74.46	41.88	46.79	40.76
R Squared	0.88	0.90	0.90	0.91	0.90	0.92	0.93	0.94
N School-Grade-Year	2691	2688	2693	2686	1571	1576	1572	1569
N School-Year	1143	1143	1143	1143	935	935	935	935
N School	451	451	451	451	379	379	379	379
Panel B. "Difference-in-Differences" Estimates								
SDrillGroup x DrillExposureIntensity	-0.24	-0.40*	-0.42**	-0.35	-0.21	-0.33	-0.50**	-0.50*
	(0.20)	(0.22)	(0.18)	(0.23)	(0.14)	(0.30)	(0.23)	(0.26)
Mean Proficiency Rate Fire Drill Exposure	70.34	37.40	52.67	38.52	74.36	43.81	47.97	43.13
R Squared	0.88	0.91	0.90	0.92	0.91	0.93	0.93	0.94
N School-Grade-Year	2466	2458	2473	2472	1454	1446	1451	1452
N School-Year	1034	1031	1036	1036	860	855	858	860

N School	439	438	438	439	386	386	387	387
Panel C. "Difference-in-Differences" Estimates Conditional on Additional Right Hand-side Terms								
SDrillGroup x $DrillExposureIntensity^2$	0.02	0.02	0.00	0.02	-0.03***	-0.04**	-0.04**	-0.03***
	(0.04)	(0.06)	(0.05)	(0.06)	(0.01)	(0.02)	(0.01)	(0.01)
Mean Proficiency Rate Fire Drill Exposure	70.35	37.4	52.66	38.51	74.36	43.81	47.97	43.13
R Squared	0.88	0.91	0.90	0.91	0.91	0.93	0.93	0.94
N School-Grade-Year	2466	2458	2473	2458	1454	1446	1451	1452
N School-Year	1034	1031	1036	1036	860	855	858	860
N School	439	438	438	439	386	386	387	387

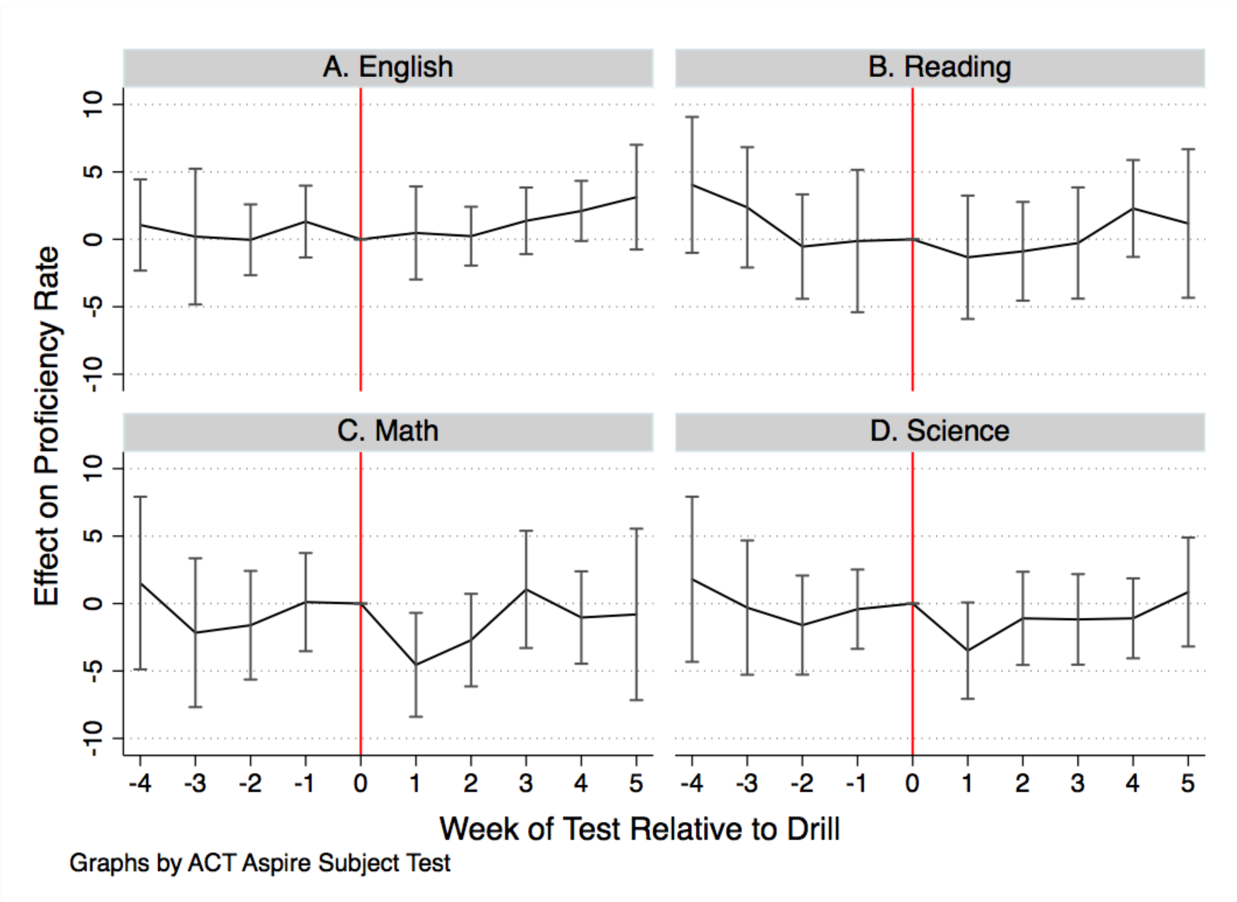
Note: The unit of observation is the school-grade-year. Each cell corresponds to a separate regression, where the dependent variable is the proficiency rate for the grade levels and subject indicated in the headers. Panel (A) corresponds to event-study model estimates recovered using Equation (2). Panel (B) corresponds to what amount to the difference-in-difference style model expressed in Equation (3). Panel (C) presents estimates recovered from Equation (3) that also include $DrillExposureIntensity^2$ and *SDrillGroup x* $DrillExposureIntensity^2$, which account for the possibility that the drill exposure effect does not fade out in a linear pattern. All specifications control for: school-level percent economically disadvantaged students; the school-grade percents for Hispanic and for Black students; the data quality indicator that denotes whether the test dates were precise to the level of the school-grade-test-year; student-teacher ratio; school-by-grade fixed effects; and year fixed effects. Estimates are weighted by the number of students in the school-grade-year taking the test. Standard errors are clustered at the school. Asterisks denote statistical significance: *p<0.1 **p<.05 ***p<.01.
Source: Arkansas Department of Education Data Center

Panel A of Table 1.4 reports the estimated effect of active-shooter drill exposure on ACT Aspire proficiency rates in English, reading, math, and science. Each cell reports the coefficient estimate from Equation (2) for the *TestWeek* indicator denoting testing the week after the drill; the reference *TestWeek* indicator denotes testing the week preceding the drill. I report coefficient estimates separately for grades 3-5 and grades 6-8. I do this because the ACT Aspire Technical Manual describes test performance and validity in these elementary and middle school grade groupings (ACT Research Services, 2019).

Columns 1-4 present estimates on the *TestWeek* indicator denoting testing in a specific subject the week after the drill for grades 3-5. For each subject test, the proficiency rate associated with testing the week after an active-shooter drill is no better or worse than it would have been had testing occurred the week leading up to the drill. That is implied by the statistically insignificant coefficient estimates in each column.

Columns 5-8 report the subject-specific estimate on the *TestWeek* indicator denoting testing the week after the drill for grades 6-8. The effect of testing the week after the active-shooter drill has effects on proficiency rates that vary by subject. Columns 5 and 6 imply that English and reading proficiency rates associated with testing the week after the drill are indistinguishable from those associated with testing the week before the drill. Column 7 implies that the math proficiency rate associated with testing the week after an active-shooter drill is 4.5 pp lower than the math proficiency rate associated with testing the week before the drill. For context, that is about 10 percent of the mean proficiency rate for middle school students testing in math the week before the active-shooter drill. Similarly, Column 8 implies that the science proficiency rate associated with testing the week after an active-shooter drill is about 3.5 pp lower than the rate associated with testing the week before the drill. That is about 9 percent of the mean proficiency rate for middle-school students testing in science the week before the drill. These declines in math and science are comparable in magnitude to the negative effect on elementary school math proficiency rates of true active-shooter threat exposure in the form of the Beltway Sniper Attacks, that is, 2-5 pp (Gershenson and Tekin 2018).

Figure 1.3: Active-shooter drill exposure and ACT Aspire proficiency rates, Arkansas grades 6-8, 2016-2019.



Notes: Each panel presents coefficient estimates (90% CIs) for the *TestWeek* indicators in the school-by-grade fixed effects specification; the *TestWeek* reference indicator equal to 0 denotes testing the week leading up to the drill. School-grade-year observations: English 1571, reading 1576, math 1572, science 1569.

Figure 1.3 presents the effects of active-shooter drill exposure on middle school proficiency rates when drill exposure intensity is weaker because the drill occurs further away in time relative to the test. This figure plots the coefficient estimates on the *TestWeek* indicator variables corresponding to the 4-5 weeks immediately before and after the drill. The reference category is the effect of active-shooter drill exposure on the subject proficiency rate associated with testing during the week preceding the drill. Exposure during the weeks surrounding the English and reading tests has no discernible effect on subject proficiency rates. However, there is a negative effect on proficiency rates in math and science when testing the week after the drill.

To summarize, there are three takeaways from the event study analysis. First, active-shooter drill exposure during the week leading up to the test appears to affect test performance among middle school but not elementary school students. Future research should seek explanations. Perhaps elementary school administrators make the drills less realistic for elementary school students, given the developmental age of these students. Indeed, the ALICE Training Institute offers age-appropriate, discussion-based protocols for discussing with K-5 students how they might respond to a "dangerous person" (Jonson, Moon, and Gialopsos 2020).

Second, the effect of active-shooter drill exposure during the week leading up to the test is fleeting, at least in math and science for middle school students. This particular result begins to undermine the claim that school active-shooter drill exposure traumatizes students. Perhaps students instead experience stress and emotional discomfort in the short run.

Finally, prior studies offer plausible explanations for why active-shooter drill exposure during the week leading up to the test influences proficiency rates differently by subject. Math skills are more sensitive than reading skills to school-environment shocks (Hanushek and Rivkin 2010). Perhaps math and science are subjects more amenable to cram studying than English and reading. Moreover, students tend to develop language and reading skills at home and math skills at school (Currie and Thomas 2000). These facts begin to explain how drill exposure during the week leading up to the test affects both math and science—two subjects related to one another in conceptual thinking and problem-solving skills—but not English and reading.

While the event-study model suggests that active-shooter drill exposure has a large negative effect for middle school math and science only, the DD-style model I use, which leverages fire drill exposure as the counterfactual condition, allows me to factor out instructional time loss and declines in student learning attributable to safety drills. Table A.3 reports summary statistics for the DD analytic sample. Column 1 reports means for the full analytic sample. Columns 2 and 3 report means for the active-shooter drill exposure group and fire drill exposure group, respectively. Column 4 reports the difference in means between the two groups as well as the result of a two-tailed t-test of the difference in means. The two groups are significantly different on most school-grade and school-level characteristics.

Panel B of Table 1.4 reports the β_3 coefficient estimate from Equation (3) by grade span and ACT Aspire subject test. There are two takeaways. First, the coefficient estimates in Columns 1-4 resemble in magnitude those recovered from the event study model, although for reading and math the estimates are significant whereas their event-study analogs are indistinguishable from zero. Although reading and math are of primary importance to federal school accountability policy and therefore the main targets of intensive and focused instruction in general and especially around test time, performance on these tests might be particularly sensitive to the unique effects of active-shooter drill exposure.

Second, the coefficient estimates reported in Columns 5-8, which correspond to grades 6-8, are all noticeably smaller in magnitude than their event-study analogs, the same in direction, and significant at conventional levels. Each point estimate ranges from one-fifth to one-half of 1 percentage point. For context, I interpret the coefficient estimate in Column 7. This estimate implies that a one-week increase in drill exposure intensity for the active-shooter drill exposure group taking the ACT Aspire Math Test is associated with a 0.50 pp decline in the math proficiency rate, relative to a one-week increase in drill exposure intensity among the fire drill exposure group. That point estimate can be contextualized in at least two ways. The first is that that amounts to about 1 percent of the fire drill exposure group's mean proficiency rate. The second is to think of the point estimate in terms of a count of students. Over the 2016-19 school years, I observe test schedules affecting an average of about 86,017 students per year in grades 6-8. Therefore, about 430 students (.50% x 86,017) were exposed to an active-shooter drill prior to the test and scored below proficient on the ACT Aspire Math Test.

Panel C of Table 1.4 reports coefficient estimates from a variant of Equation (3) in which the right hand-side includes $DrillExposureIntensity^2$ and $SDrillGroup \times DrillExposureIntensity^2$. While measured with precision, the coefficient estimate on the interaction term is effectively zero (elementary school) or quite modest in magnitude. Thus, it appears appropriate to model the effect of active-shooter drill exposure on proficiency rates as a simple linear function.

Table 1.5 reports coefficient estimates from Equation (3) after the exclusion of any school-grade-subject-year observation for which the active-shooter drill and the fire drill occurred within 7, 21, or 35 days of one another. I restrict this analysis to reading and math for grades 3-5 and math and science for grades 6-8. I do this because the DD estimates reported in Panel B of Table 1.4 were significant only for these grade-test combinations.

Table 1.5: "Difference-in-differences" alternative specifications for the effect of active-shooter drill exposure on proficiency rates, by grade level groupings and ACT Aspire Test subject, Arkansas Public Schools, 2016-2019.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Grades 3-5						Grades 6-8					
	Reading			Math			Math			Science		
	7 days	21 days	35 days	7 days	21 days	35 days	7 days	21 days	35 days	7 days	21 days	35 days
Panel A. Re-estimate Equation (3), Drop Close Cases												
SDrillGroup x DrillExposureIntensity	-0.17	-0.19	0.06	-0.57**	-0.74	-1.11	-0.76**	-0.94	-1.12**	-0.85**	-0.44	-0.53
	(0.26)	(0.41)	(0.80)	(0.27)	(0.47)	(0.68)	(0.38)	(0.67)	(0.56)	(0.36)	(0.80)	(0.97)
R Squared	0.91	0.92	0.92	0.91	0.92	0.93	0.93	0.94	0.94	0.94	0.95	0.96
N School-Grade-Year	2329	2094	1864	2349	2112	1885	1348	1145	1029	1345	1139	1027
N School-Year	976	876	780	983	880	786	804	690	616	803	687	615
N School	433	420	393	433	420	394	379	361	338	379	360	337
Panel B. Re-estimate Equation (3), Prioritize Active-Shooter Drills												
SDrillGroup x DrillExposure Intensity	-0.33	-0.36*	-0.27	-0.37*	-0.39**	-0.20	-0.50**	-0.46***	-0.44***	-0.55**	-0.47**	-0.46**
	(0.24)	(0.20)	(0.20)	(0.20)	(0.18)	(0.19)	(0.20)	(0.15)	(0.15)	(0.23)	(0.19)	(0.18)
R Squared	0.91	0.91	0.91	0.90	0.90	0.90	0.93	0.93	0.93	0.94	0.94	0.94
N School-Grade-Year	2458	2458	2458	2473	2473	2473	1451	1451	1451	1452	1452	1452
N School-Year	1031	1031	1031	1036	1036	1036	858	858	858	860	860	860
N School	438	438	438	438	438	438	387	387	387	387	387	387

Notes: The unit of observation is the school-grade-year. Each cell corresponds to a separate regression, where the dependent variable is the proficiency rate for the subject indicated in the column header. Panel (A) presents results from re-estimation of Equation (3) in which school-grade-subject-year observations for which an active-shooter drill and fire drill occurred within 7, 21, or 35 days of one another are dropped from the analytic sample. Panel (B) presents results from re-estimation of Equation (3) in which school-grade-subject-year observations for which an active-shooter drill and fire drill occurred within 7, 21, or 35 days of one another are classified into the active-shooter drill exposure group. All specifications control for: school-level percent economically disadvantaged students; the school-grade percents for Hispanic and for Black students; the data quality indicator that denotes whether the test dates were precise to the level of the school-grade-test-year; student-teacher ratio; school-by-grade fixed effects; and year fixed effects. Estimates are weighted by the number of students in the school-grade-year taking the test. Standard errors are clustered at the school. Asterisks denote statistical significance: *p<0.1 **p<.05 ***p<.01.
Source: Arkansas Department of Eduation Data Center

Panel A of Table 1.5 presents coefficient estimates recovered from Equation (3) after restricting the analytic sample only to fire and active-shooter drills whose effects are likely to operate on proficiency rates independently of one another. These estimates deviate from their counterparts in Panel B of Table 1.4 in ways that are consistent with student performance on math tests being more sensitive to lost instructional time and lowered student learning than student performance on reading tests. Columns 1-3 present reading estimates that are no longer significant, and that are noticeably smaller than their counterpart in Column 2 in Panel B of Table 1.4. Yet in Column 4 the math estimate is much larger in magnitude relative to its Table 1.5 counterpart (Column 3, Panel B), although the estimates in Columns 5 and 6 are insignificant. Columns 7-9 document a much clearer pattern in math proficiency rates for grades 6-8, which suggests that the unique effect of active-shooter drill exposure on middle school student performance on math tests might be larger than initially estimated. Columns 10-12 document a mixed pattern in science proficiency rates for grades 6-8.

Panel B of Table 1.5 presents coefficient estimates recovered from Equation (3) after classifying certain school-grade-subject-year observations into the active-shooter drill exposure group instead of the fire drill exposure group. These estimates are qualitatively similar to their counterparts in Panel B of Table 1.4, implying that these particular observations do not drive the main results. The takeaway here is that the rather simple classification scheme I presented initially (i.e., the classification of observations into the active-shooter drill exposure group or the fire drill exposure groups based on temporal closeness to the test) may be quite reasonable.

1.6.4 Robustness Checks

Appendix 2 presents the results of robustness checks in which I re-estimate Equations (1)-(3) over analytic samples that exclude schools reporting active-shooter drills lasting zero minutes in 2016. These checks assess the extent to which these particular schools drive the main results reported in Tables 1.3 and 1.4. I obtain estimates similar to those in Tables 1.3 and 1.4.

1.6.5 Falsification Tests

I conduct falsification tests to address the objection that the timing of the current year's active-shooter drill could be determined by the prior year's proficiency rates. Perhaps principals suspect that active-shooter drills occurring prior to the test have a negative effect on proficiency rates. Therefore, principals schedule these drills further away in time from the ACT Aspire Tests in the subsequent year. To assess this possibility, I regress the *prior year t*-1 proficiency rate on three separate indicator variables denoting the *current year t* timing of the active-shooter drill relative to the test: (a) before the school ACT Aspire testing window, (b) before the school window but during the fall semester, and (c) after the school ACT Aspire testing window. I choose these three test-drill timings because principals might only be able to schedule these drills within relatively large segments of the academic calendar rather than exact weeks. Moreover, because federal school accountability policies raise the stakes of math and reading scores, I assume the proficiency rates for these two subject tests would influence principals' decisions.

Table 1.6: Falsification test regressing prior year ACT proficiency rate on indicator for timing of active-shooter drill relative to test, Arkansas Public Schools, 2016-2019.

	(1)	(2)	(3)
	Drill Relative to Test		
	Before School ACT Window		
	Drill Anytime	Drill Fall Semester	Anytime After School ACT Window
Panel A. Once-Lagged Math Proficiency Rate			
Coefficient Estimate	0.103	0.863	0.034
	(0.946)	(0.743)	(0.991)
R Squared	0.908	0.908	0.908
N School-Grade-Year	3479	3479	3479
N School-Grade	1747	1747	1747
N School	638	638	638
Panel B. Once-Lagged Reading Proficiency Rate			
Coefficient Estimate	-0.224	0.522	0.423
	(0.957)	(0.572)	(1.017)
R Squared	0.918	0.918	0.918
N School-Grade-Year	3479	3479	3479
N School-Grade	1747	1747	1747
N School	638	638	638

Note: Panel (A) reports the coefficient estimate on the once-lagged math proficiency rate. Panel (B) reports the coefficient estimate on the once-lagged reading proficiency rate. All models control for: school-by-grade and year fixed effects, school-level percent economically disadvantaged (equivently the share of students receiving free or reduced-price lunch); school-grade percents of Hispanic and of Black students; and the current year's data quality indicator for whether the recovered test dates are precise to the level of the school-grade-subject. Estimates are weighted by the number of students in the school-grade-test-year cell. Standard errors are clustered on the school. Asterisks denote statistical significance: *p<0.1 **p<.05 ***p<.01.
Source: Arkansas Department of Eduation Data Center

Table 1.6 reports the results of these falsification tests. Every point estimate is statistically indistinguishable from zero and imprecise. Therefore, the current year's test-drill temporal proximity does not appear to predict the prior year's math and reading proficiency rates.

1.7 Discussion

1.7.1 Limitations

First, there are drawbacks to the use of percent proficient metrics as outcome measures. Proficiency thresholds, or the cut scores separating one score category from another (e.g., Proficient versus Distinguished) are effectively arbitrary. Both trend and gap magnitudes for the proficiency rate vary systematically with the positions of these thresholds (Ho 2008; Lu 2017). Moreover, in this study, these metrics capture changes in achievement only for the marginally non-proficient (or proficient) student whose test performance responds unfavorably to active-shooter drill exposure. The use of percent proficient metrics preclude researchers from investigating changes in achievement for students who typically score further "away" from the thresholds and for whom active-shooter drill exposure would move their scores but within the same score category (e.g., low- to mid-proficient) (Jacob et al. 2014). Future studies might address this limitation by drawing upon deidentified, student-level scale scores for tests, and collapsing these data to the desired aggregate level (e.g., school-grade-year cell).

Second, I am unable to directly assess the credibility of the Trauma Mechanism. One way to do so could be to use measures of student well-being as a stand-in for trauma. Student well-being is the "psychological, cognitive, social and physical functioning and capabilities that students need to live a happy and fulfilling life" (OECD 2017, p. 61), which includes personal security and subjective well-being (OECD 2017). Privacy policies restrict what researchers can observe about the physical or social-emotional health of minors. Researchers might instead consider proxies for student well-being. For example, Rossin-Slater et al. (2020) document a significant and positive effect of school shootings on the post-shooting volume of antidepressant prescriptions in nearby geographic areas, relative to areas located farther away from such shootings. Another potentially fruitful data source might be counts of mental health emergency department visits for children under 18 (Leeb et al. 2020). With a sufficiently large sample, a researcher might aggregate counts over fine geographic areas and time periods to proxy for student well-being around the time of school active-shooter drills.

1.7.2 Conclusions

While active-shooter drills are likely to remain an essential safety practice in schools, school administrators can take straightforward steps to mitigate the negative effects I documented in the current study. One solution is to identify students who would have especially unfavorable emotional responses to these drills and offer them the option to engage in less intense discussion-based protocols. Principals can also avoid scheduling these drills around test dates and instead within the first 1-2 months of the school year. This scheduling strategy would also ensure that most students, for the duration of the school year, are aware of the response protocols for active-shooter threats. Yet another solution, albeit more demanding, would require the close examination of, and perhaps modification to, how these drills are conducted. Some argue that active-shooter drills are “likely to cause significant distress and psychological harm” if they involve deception and even if the drills are *announced*, that is, whether students and staff are privy to the drill in advance (Schonfeld, Rossen, and Woodard 2017). Classroom discussions throughout the year about active-shooter incidents reported in the media and press might be one way to acclimate students to the seriousness of active-shooter treats and drills.

REFERENCES

Abouk, Rahi, and Scott Adams. 2013. "School Shootings and Private School Enrollment." *Economics Letters* 118 (2): 297–99. https://doi.org/10.1016/j.econlet.2012.11.009.

ACT. 2018. "Interpretive Guide for ACT Aspire Summative Reports." ACT, Inc. http://avocet.pearson.com/ArkansasACTAspire/Home#pdf=7925_17488#page=1.

ACT Research Services. 2019. "ACT Aspire Summative Technical Manual, Ver. 5." Iowa: ACT, Inc. https://www.act.org/content/dam/act/unsecured/documents/2019/aspire/Aspire-Summative-Technical-Manual.pdf.

ALICE Training. 2020. "Active Shooter Preparedness | ALICE Violent Intruder Safety Training Program." *ALICE Training* (blog). November 9, 2020. https://www.alicetraining.com/our-program/.

ALICE Training Institute. 2020. "ALICE: An Easy to Remember Acronym." 2020. https://www.alicetraining.com/our-program/alice-training/.

Allen, Jeff, and Justine Radzunel. 2017. "What Are the ACT College Readiness Benchmarks?" Issue Brief. https://www.act.org/content/dam/act/unsecured/documents/pdfs/R1670-college-readiness-benchmarks-2017-11.pdf.

Altmann, Erik M., and J. Gregory Trafton. 2004. "Task Interruption: Resumption Lag and the Role of Cues." *Proceedings of the Cognitive Science Society* 26 (26). https://apps.dtic.mil/sti/pdfs/ADA480333.pdf.

Altmann, Erik M., J. Gregory Trafton, and David Z. Hambrick. 2014. "Momentary Interruptions Can Derail the Train of Thought." *Journal of Experimental Psychology: General* 143 (1): 215–26. https://doi.org/10.1037/a0030986.

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders*. Washington, D.C.: American Psychiatric Association.

Aucejo, Esteban M., and Teresa Foy Romano. 2016. "Assessing the Effect of School Days and Absences on Test Score Performance." *Economics of Education Review* 55 (December): 70–87. https://doi.org/10.1016/j.econedurev.2016.08.007.

Beland, Louis-Philippe, and Dongwoo Kim. 2016. "The Effect of High School Shootings on Schools and Student Performance." *Educational Evaluation and Policy Analysis* 38 (1): 113–26. https://doi.org/10.3102/0162373715590683.

Christakis, Erika. 2019. "Active-Shooter Drills Are Tragically Misguided." *The Atlantic*, March 2019. https://www.theatlantic.com/magazine/archive/2019/03/active-shooter-drills-erika-christakis/580426/.

Currie, Janet, and Duncan Thomas. 2000. "School Quality and the Longer-Term Effects of Head Start." *The Journal of Human Resources* 35 (4): 755. https://doi.org/10.2307/146372.

Diliberti, Melissa, Michael Jackson, Samuel Correa, and Zoe Padgett. 2019. "Crime, Violence, Discipline, and Safety in US Public Schools: Findings from the School Survey on Crime and Safety: 2017-18. First Look. NCES 2019-061." NCES 2019-061. Washington, D.C.: National Center for Education Statistics. https://files.eric.ed.gov/fulltext/ED596638.pdf.

Diliberti, Melissa, Michael Jackson, and Jana Kemp. 2017. "Crime, Violence, Discipline, and Safety in US Public Schools: Findings from the School Survey on Crime and Safety: 2015-16. First Look. NCES 2017-122." NCES 2017-122. Washington, D.C.: National Center for Education Statistics. https://files.eric.ed.gov/fulltext/ED574956.pdf.

Donovan, Colleen, David N. Figlio, and Mark Rush. 2006. "Cramming: The Effects of School Accountability on College-Bound Students." National Bureau of Economic Research.

Everytown for Gun Safety Support Fund. 2020. "The Impact of School Safety Drills for Active Shootings." everytownresearch.org/school-safety-drills.

Federal Bureau of Investigation. 2019. "Quick Look: 277 Active Shooter Incidents in the United States From 2000 to 2018." Washington, D.C.: U.S. Department of Justice. https://www.fbi.gov/about/partnerships/office-of-partner-engagement/active-shooter-incidents-graphics.

Foerde, Karin, Barbara J. Knowlton, and Russell A. Poldrack. 2006. "Modulation of Competing Memory Systems by Distraction." *Proceedings of the National Academy of Sciences* 103 (31): 11778–83. https://doi.org/10.1073/pnas.0602659103.

Gershenson, Seth, and Erdal Tekin. 2018. "The Effect of Community Traumatic Events on Student Achievement: Evidence from the Beltway Sniper Attacks." *Education Finance and Policy* 13 (4): 513–44. https://doi.org/10.1162/edfp_a_00234.

Gillie, Tony, and Donald Broadbent. 1989. "What Makes Interruptions Disruptive? A Study of Length, Similarity, and Complexity." *Psychological Research* 50 (4): 243–50. https://doi.org/10.1007/BF00309260.

Goodman, Joshua. 2014. "Flaking out: Student Absences and Snow Days as Disruptions of Instructional Time." National Bureau of Economic Research.

Guerino, Paul, Michael D. Hurwitz, Margaret E. Noonan, and Sarah M. Kaffenberger. 2006. "Crime, Violence, Discipline, and Safety in US Public Schools: Findings from the School Survey on Crime and Safety: 2003-04. First Look. NCES 2007-302." NCES 2007-302. Washington, D.C.: National Center for Education Statistics. https://files.eric.ed.gov/fulltext/ED494758.pdf.

Hanushek, Eric A, and Steven G Rivkin. 2010. "Generalizations about Using Value-Added Measures of Teacher Quality." *American Economic Review* 100 (2): 267–71. https://doi.org/10.1257/aer.100.2.267.

Ho, Andrew Dean. 2008. "The Problem With 'Proficiency': Limitations of Statistics and Policy Under No Child Left Behind." *Educational Researcher* 37 (6): 351–60. https://doi.org/10.3102/0013189X08323842.

Huskey, Michael G., and Nadine M. Connell. 2020. "Preparation or Provocation? Student Perceptions of Active Shooter Drills." *Criminal Justice Policy Review*, January, 088740341990031. https://doi.org/10.1177/0887403419900316.

Jacob, Robin T., Roger D. Goddard, and Eun Sook Kim. 2014. "Assessing the Use of Aggregate Data in the Evaluation of School-Based Interventions: Implications for Evaluation Research and State Policy Regarding Public-Use Data." *Educational Evaluation and Policy Analysis* 36 (1): 44–66. https://doi.org/10.3102/0162373713485814.

Jonson, Cheryl Lero. 2017. "Preventing School Shootings: The Effectiveness of Safety Measures." *Victims & Offenders* 12 (6): 956–73. https://doi.org/10.1080/15564886.2017.1307293.

Jonson, Cheryl Lero, Melissa M. Moon, and Brooke Miller Gialopsos. 2020. "Are Students Scared or Prepared? Psychological Impacts of a Multi-Option Active Assailant Protocol Compared to Other Crisis/Emergency Preparedness Practices." *Victims & Offenders; Philadelphia* 15 (5): 639–62. http://dx.doi.org.libproxy.lib.unc.edu/10.1080/15564886.2020.1753871.

Jonson, Cheryl Lero, Melissa M. Moon, and Joseph A. Hendry. 2020. "One Size Does Not Fit All: Traditional Lockdown Versus Multioption Responses to School Shootings." *Journal of School Violence* 19 (2): 154–66. https://doi.org/10.1080/15388220.2018.1553719.

Kamenetz, Anya. 2019. "A Look At The Value Of Active Shooter Drills In Schools." *National Public Radio*, November 22, 2019. https://www.npr.org/2019/11/22/782130813/a-look-at-the-value-of-active-shooter-drills-in-schools.

Kamentz, Anya. 2020. "2 Big Teachers Unions Call For Rethinking Student Involvement In Lockdown Drills." *National Public Radio*, February 11, 2020. https://www.npr.org/2019/11/22/782130813/a-look-at-the-value-of-active-shooter-drills-in-schools.

Kraft, Matthew A., and Manuel Monti-Nussbaum. 2020. "The Big Problem with Little Interruptions to Classroom Learning." *EdWorkingPapers.Com*. Annenberg Institute at Brown University. https://www.edworkingpapers.com/ai20-227.

Laurito, Agustina, Johanna Lacoe, Amy Ellen Schwartz, Patrick Sharkey, and Ingrid Gould Ellen. 2019. "School Climate and the Impact of Neighborhood Crime on Test Scores." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5 (2): 141–66. https://doi.org/10.7758/RSF.2019.5.2.08.

Leeb, Rebecca T. 2020. "Mental Health–Related Emergency Department Visits Among Children Aged 18 Years During the COVID-19 Pandemic — United States, January 1–October 17, 2020." *MMWR. Morbidity and Mortality Weekly Report* 69. https://doi.org/10.15585/mmwr.mm6945a3.

Liu, Jing, Monica Lee, and Seth Gershenson. 2019. "The Short- and Long-Run Impacts of Secondary School Absences." SSRN Scholarly Paper ID 3457657. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=3457657.

Lu, Ying. 2017. "Variability in Percentage Above Cut Scores Due to Discreteness in Score Scale." *ETS Research Report Series* 2017 (1): 1–11. https://doi.org/10.1002/ets2.12158.

Marcotte, Dave E., and Steven W. Hemelt. 2008. "Unscheduled School Closings and Student Performance." *Education Finance and Policy* 3 (3): 316–38. https://doi.org/10.1162/edfp.2008.3.3.316.

Musu, L., A. Zhang, K. Wang, J. Zhang, and B. A. Oudekerk. 2019. "Indicators of School Crime and Safety." 2019-047 252571. Washington, D.C.: United States Department of Education National Center for Education Statistics. https://files.eric.ed.gov/fulltext/ED594245.pdf.

NASP. 2017. "Best Practice Considerations for Schools in Active Shooter and Other Armed Assailant Drills." National Association of School Psychologists. https://www.nasponline.org/Documents/Research%20and%20Policy/Advocacy%20Resources/BP_Armed_Assailant_Drills.pdf.

NASP. 2018. "Mitigating Negative Psychological Effects of School Lockdowns: Brief Guidance for Schools." National Association of School Psychologists. https://www.nasponline.org/Documents/Resources%20and%20Publications/Resources/Crisis/Mitigating_Psychological_Effects_of_Lockdowns.pdf.

Neiman, Samantha. 2011. "Crime, Violence, Discipline, and Safety in US Public Schools. Findings from the School Survey on Crime and Safety: 2009-10. First Look. NCES 2011-320." NCES 2011-320. Washington, D.C.: National Center for Education Statistics. https://files.eric.ed.gov/fulltext/ED520120.pdf.

Neiman, Samantha, and Jill F. DeVoe. 2009. "Crime, Violence, Discipline, and Safety in US Public Schools: Findings from the School Survey on Crime and Safety, 2007-08." National Center for Education Statistics, Institute of Education Sciences ….

Nolle, Kacey Lee, Paul Guerino, Rachel Dinkes, and Kathryn Chandler. 2007. "Crime, Violence, Discipline, and Safety in US Public Schools. Findings from the School Survey on Crime and Safety: 2005-06. NCES 2007-361." NCES 2007-361. Washington, D.C.: National Center for Education Statistics. https://files.eric.ed.gov/fulltext/ED498378.pdf.

OECD. 2017. "PISA 2015 Results (Volume III): Students' Well-Being." OECD Publishing. https://www.oecd-ilibrary.org/education/pisa-2015-results-volume-iii_9789264273856-en.

O'Regan, Sylvia Varnham. 2019. "The Company Behind America's Scariest School Shooter Drills." *The Trace*, December 13, 2019. https://www.thetrace.org/2019/12/alice-active-shooter-training-school-safety/.

Peterson, Jillian, Ellen Sackrison, and Angela Polland. 2015. "Training Students to Respond to Shootings on Campus: Is It Worth It?" *Journal of Threat Assessment and Management* 2 (2): 127–38. https://doi.org/10.1037/tam0000042.

Pynoos, Robert S., Calvin Frederick, Kathi Nader, William Arroyo, Alan Steinberg, Spencer Eth, Francisco Nunez, and Lynn Fairbanks. 1987. "Life Threat and Posttraumatic Stress in School-Age Children." *Archives of General Psychiatry* 44 (12): 1057–63. https://doi.org/10.1001/archpsyc.1987.01800240031005.

Richter, Ed. 2019. "'I Started to Cry': Franklin Students Describe Active Shooter Training Involving Blanks." *Dayton Daily News*, October 2, 2019. https://www.daytondailynews.com/news/local/started-cry-franklin-students-describe-active-shooter-training-involving-blanks/mT25ccvGSwLP2UdNq8mPVL/.

Rossin-Slater, Maya, Molly Schnell, Hannes Schwandt, Sam Trejo, and Lindsey Uniat. 2020. "Local Exposure to School Shootings and Youth Antidepressant Use." *Proceedings of the National Academy of Sciences*, September, 202000804. https://doi.org/10.1073/pnas.2000804117.

Schildkraut, Jaclyn, Amanda B. Nickerson, and Thomas Ristoff. 2020. "Locks, Lights, Out of Sight: Assessing Students' Perceptions of Emergency Preparedness across Multiple Lockdown Drills." *Journal of School Violence* 19 (1): 93–106. https://doi.org/10.1080/15388220.2019.1703720.

Schonfeld, David J., Eric Rossen, and Diann Woodard. 2017. "Deception in Schools—When Crisis Preparedness Efforts Go Too Far." *JAMA Pediatrics* 171 (11): 1033–34. https://doi.org/10.1001/jamapediatrics.2017.2565.

Sharkey, P. 2010. "The Acute Effect of Local Homicides on Children's Cognitive Performance." *Proceedings of the National Academy of Sciences* 107 (26): 11733–38. https://doi.org/10.1073/pnas.1000690107.

Sharkey, Patrick, Amy Ellen Schwartz, Ingrid Gould Ellen, and Johanna Lacoe. 2014. "High Stakes in the Classroom, High Stakes on the Street: The Effects of Community Violence on Student's Standardized Test Performance." *Sociological Science* 1: 199–220. https://doi.org/10.15195/v1.a14.

Sharkey, Patrick T., Nicole Tirado-Strayer, Andrew V. Papachristos, and C. Cybele Raver. 2012. "The Effect of Local Violence on Children's Attention and Impulse Control (5th Edition)." *American Journal of Public Health* 102 (12): 2287–93. https://doi.org/10.2105/AJPH.2012.300789.

Smith, Carolyn, and Bonnie E. Carlson. 1997. "Stress, Coping, and Resilience in Children and Youth." *Social Service Review* 71 (2): 231–56. https://www.jstor.org/stable/30013020.

Zhe, Elizabeth J, and Amanda B Nickerson. 2007. “Effects of an Intruder Crisis Drill on Children’s Knowledge, Anxiety, and Perceptions of School Safety.” *School Psychology Review* 36 (3): 10.

CHAPTER 2: KINDERGARTEN ENROLLMENT AGE AND HIGH-FREQUENCY

2.1 Overview

Relative to younger kindergarten entrants, do older kindergarten entrants have academic and behavioral advantages during the early grades and, importantly, within these grade levels? We exploit plausibly exogenous variation in kindergarten enrollment age resulting from a state policy that requires students to reach five-years-old by a certain date to enroll in kindergarten in that same calendar year. Students turning five-years-old after that date must enroll the following year, making them the oldest in their kindergarten cohorts. We use a regression discontinuity approach to exploit this variation in kindergarten enrollment age, and we control for rich measures of kindergarten readiness to account for human capital investments made prior to the kindergarten year. We find that older kindergarten entrants, while ahead of their younger counterparts in foundational literacy skills at the beginning and the middle of kindergarten and first grade, become similar to their younger counterparts another by the end of each year. We also find that older kindergarten entrants have little to no advantage over their younger counterparts in their capacity to cooperate with their peers during kindergarten and first grade, although teachers perceive older entrants as consistently more adept at engaging independently in schoolwork. Finally, we find that older kindergarten entrants are as likely as their younger counterparts to be identified with a disability between the start of kindergarten and the end of third grade.

2.2 Introduction

One of the earliest schooling decisions parents face is whether to *redshirt* their children, that is, to intentionally delay the child's entry into kindergarten so the child is relatively older than the child's peers (Bassok and Reardon, 2013). The prevailing wisdom behind redshirting is that it conveys upon the child advantages over the child's younger peers in academics, behaviors, social interactions, and athletics (Deming and Dynarski, 2008). By some measure, these advantages extend far beyond elementary school. For example, older kindergarten entrants are less likely classified as eligible for special education and more likely as gifted between third and eighth grade (Dhuey et al., 2019), and are more likely to enroll in post-secondary education immediately after high school graduation (Hemelt and Rosen, 2016). But relatively little is known about how, and the extent to which, the age advantage changes *within* grade levels, especially during the early elementary school years. Are there moments within each grade when the age advantage in academics and behaviors is especially large or small, or when it declines or rises noticeably relative?

Prior studies addressing relevantly similar questions faced trade-offs between methodological capabilities and substantive contributions. On the one hand, studies have drawn upon large administrative datasets stored in Statewide Longitudinal Data Systems (SLDS), which allow analysts to leverage the "preferred" (Dhuey et al. 2019) regression discontinuity (RD) approach to examine the relationship between age at kindergarten entry and student outcomes. Increasingly large volumes of student-level data in SLDS render feasible the RD approach, whose requirement for statistical power is non-trivial. However, researchers using SLDS often restrict their analytic focus to 3rd-8th grade outcomes (e.g., test scores), because school systems collect these outcomes to fulfill federal and state school accountability policies. On the other hand, studies that investigate outcomes measured in grades K-2 draw upon smaller but nationally representative datasets, such as the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K). The outcomes available in these datasets are incredibly rich, as they often include measures of cognitive and noncognitive (behavioral) outcomes that, for a number of reasons (e.g., fiscal, political), are typically unavailable in SLDS. However, the relatively small samples in nationally representative datasets force researchers to deviate methodologically from the preferred RD approach. Typically, these analysts leverage a related quasi-experimental approach (i.e., instrumental variables [IV] using expected age at kindergarten entry to instrument for actual age at kindergarten entry) to obtain plausibly causal estimates of age at kindergarten entry on student outcomes (Datar 2006; Elder and Lubotsky 2009; Lubotsky and Kaestner 2016). Moreover, we are aware of no study in this literature that has directly accounted for heterogeneous skill differences at the time of kindergarten entry.

I minimize the trade-offs that limited prior studies by implementing the preferred RD approach and drawing upon student-level administrative data containing outcomes for students in grades K-3 as well as measures of human capital existing prior to kindergarten. I estimate the causal effect of kindergarten enrollment at an older age on measures of foundational literacy skills, teacher ratings of student behaviors, and disability status (or special education [SPED] classification). I address four main research questions. First, to what extent does the age advantage in foundational literacy skills had by older kindergarten entrants grow or decline during grades K-1? Second, do teachers perceive differences in student behaviors among older and younger kindergarten entrants during grades K-1? Moreover, do these perceptions change during the year as teachers and students become more familiar with one another, and presumably as students learn to execute the expected classroom behaviors? Third, are older kindergarten entrants more or less likely than their younger counterparts to be identified with a disability during grades K-3, and does the probability of disability identification change within certain grades? Finally, to what extent do measures of kindergarten readiness affect the results?

I preview the three main results. First, older kindergarten entrants possess a meaningful advantage over younger kindergarten entrants in foundational literacy skills at the beginning and the middle of the year, in both kindergarten and first grade. At the beginning of kindergarten, for example, older entrants are about 0.4 SD above their younger peers in their foundational literacy skills, conditional on measures of kindergarten readiness. Reassuringly, this particular estimate we recover is consistent with a previously reported same-period estimate of the effect of older age at kindergarten enrollment on measures of reading achievement in the ECLS-K (Elder and Lubotsky 2009). However, the age advantage in foundational literacy skills fades out rapidly diminishes during kindergarten and again during first grade, and by the end of each grade it vanishes. This result complements recent findings suggesting that only after first grade do older kindergarten entrants lose any such advantage in measures of reading and literacy (Lubotsky and Kaestner 2016).

Second, while teachers perceive older and younger kindergarten entrants as similarly skilled at cooperating with peers, teachers consistently perceive younger entrants as less capable than their older counterparts in completing tasks independently, at every quarter from the start of kindergarten through the end of first grade. This particular result is notable, given how much more skilled at completing independent work one might expect even the youngest kindergarten entrant to become during kindergarten and first grade. The familiarity that teachers have with their students would suggest these teachers are attuned to their students’ improvements and setbacks, particularly in behavioral outcomes.

Finally, through the end of third grade, older kindergarten entrants are generally less likely than younger entrants to be identified with any disability. This result is somewhat inconsistent with Dhuey et al. (2019)'s recent findings from Florida showing that, during grades 3-8, older kindergarten entrants are 3-5 percentage points less likely than their younger counterparts to be identified as having any disability. However, consistent with Dhuey et al. (2019) is that older kindergarten entrants in my study are as likely as younger entrants to be identified as having a cognitive disability.

This study makes substantive extensions to prior studies that examine the relationship between kindergarten enrollment age and early elementary school outcomes. It does so by drawing upon a rich collection of student outcomes. First, I extend prior studies whose measures of reading achievement are commonly available in SLDS (e.g., 3^{rd}-8^{th} end-of-grade standardized tests) or in the ECLS-K but only in the fall and spring of kindergarten and grades 1, 3, and 5. Notably, foundational literacy skills as the primary cognitive outcome yields insight into developmental differences between older and younger kindergarten entrants. I think of these differences as "first order" differences, inasmuch as they have a critical bearing on future reading achievement. Second, by using teachers' perceptions of student behaviors, I complement Lubotsky and Kaestner (2016), whose measures of noncognitive outcomes are based on psychometrically validated instruments.[3] Third, I am aware of only two studies that apply the RD approach to estimate the effect of older-age kindergarten enrollment on disability status in Florida (Dhuey et al. 2019) and Michigan (Shapiro 2020), collectively over grades K-8. This study's use of disability status from one of the largest school districts in the country adds to the contexts in which policymakers can understand the effects of older-age kindergarten enrollment

[3] See Tourangeau et al. (2009) for descriptions of cognitive and noncognitive skills available in the ECLS-K.

on the demand for special-education services. Finally, that this study's outcomes are measured at least three times per school year allows me to look for within-grade (i.e., mid-year) gaps between older and younger kindergarten entrants during the early elementary grades.

This study also adds to the collective understanding of skill-based complementarities, especially regarding the importance of early human capital investments for similar or dissimilar future skills. Between- and within-skill complementarity implies that the returns to investments in skill are greater for individuals with higher initial levels of different or same skills, respectively (see Cunha and Heckman 2007; Cunha, Heckman, and Schennach 2010). Elder and Lubotsky (2009) argued that skills accumulated prior to kindergarten drive early elementary school outcomes, based only on the large gap between older and younger kindergarten entrants at the start of kindergarten. Lubotsky and Kaestner (2016) also found that older kindergarten entrants scored higher on measures of cognitive and noncognitive skills at the start of kindergarten and incurred larger gains in reading and math than younger entrants. To extend these two studies, I directly control for measures of kindergarten readiness, which allows me to account for heterogeneous human capital investments made during the years of one's life preceding entrance into the formal K-12 schooling system.

This paper is organized in the following way. In Section 1, I discuss the data and measures I use in my study. In section 2, I lay out the empirical strategy. I report the results in Section 3 and conclude in Section 4 with a discussion of the findings and their broader implications.

2.3 Data

For all analyses, I use student-level administrative data collected by the Wake County Public School System (WCPSS) in North Carolina, the largest school district in the state in terms of student enrollment and among the 20 largest districts in the country on this metric.

Collectively, the data in my study data span the 2006-2019 school years,[4] and include measures of foundational literacy skills, student behaviors, disability status, and kindergarten readiness. To my knowledge, no study on kindergarten enrollment age and early elementary school outcomes has drawn upon an administrative dataset containing such a rich set of student information.

I investigate the effects of older kindergarten enrollment age on three types of outcome. The first is a measure of foundational literacy skills from a universal screener called the Dynamic Indicators of Basic Early Literacy (DIBELS; Next Edition). Teachers administer universal screeners periodically during the school year to identify students who may be at risk for developing learning difficulties and thus the students who should receive supplemental instruction or intervention (Fuchs et al., 2007; Jenkins et al., 2007). During grades K-3, teachers administer DIBELS to each student three times per year (at beginning-, middle-, and end-of-year) to assess the student's foundational literacy skills. These skills are assessed on subtests of first sound fluency, letter naming fluency, non-sense word fluency, phonemic segmentation, oral reading fluency, and story re-tell fluency. The subtests constituting the composite score change across benchmark periods, reflecting the idea that children develop facility with certain literacy skills around certain grade-periods.[5] By allowing the teacher to periodically "benchmark" (i.e., compare) the student's foundational literacy skill levels against expected levels, the teacher can identify students at risk of falling behind in their development of such skills and adjust instruction accordingly. The specific DIBELS outcome variable I use is the composite score at

[4] I refer to the school year using the year of the spring semester, e.g., 2014 refers to the 2013-14 school year.

[5] See the full timeline detailing subtest administration: https://dibels.uoregon.edu/docs/marketplace/dibels/DIBELS-Next-Administration-Timeline.pdf.

every grade-benchmark period in kindergarten and first grade. I standardize the composite score by year, grade, and benchmark period to have mean 0 and unit standard deviation.

The second type of outcome is teacher ratings of student behaviors. Specifically, there are two such outcomes, Conduct and Work Habits, which the teacher assesses four times per year, or once for each quarterly, or nine-week, report card.[6] Conduct refers to how much the student cooperates with others, respects others, and observes rules and procedures. Work Habits refers to how much the student uses time wisely, listens carefully, completes assignments, writes legibly, works independently or seeks help when needed, and completes work. Conduct and Work Habits are each rated on the following three-point scale: *Does Not Meet Expectations*, *Inconsistently Meets Expectations*, and *Meets Expectations*. Across all student-quarter observations, the modal rating for each Conduct and Work Habits is *Meets Expectations*. Therefore, I derive a Conduct indicator variable and a Work Habits indicator variable that each equals 1 if the rating is either *Does Not Meet Expectations* or *Inconsistently Meets Expectations*, 0 if *Meets Expectations*. Therefore, each such variable conveys information about behavioral deficits based on age at kindergarten enrollment. The frequency of these behavioral outcomes and the constructs they measure complement noncognitive outcomes available in the ECLS-K, namely Approaches to Learning (analogous to Work Habits) and Interpersonal Skills (analogous to Conduct),[7] which, again, are only available during the fall and spring of kindergarten and first grade.

[6] See https://www.wcpss.net/Page/1886. To the best of my knowledge, these are "in-house" outcomes created by the district.

[7] According to descriptions from the ECLS-K manual, Approaches to Learning "[m]easures behaviors that afffect the ease with which children can benefit from the learning environment. It includes six items that rate the child's attentiveness, task persistence, eager- ness to learn, learning independence, flexibility, and organization." Interpersonal Skills is a five-item scale that rates "the child's skill in forming and maintaining friendships, getting along with people who are different, comforting or helping other children, expressing feelings, ideas and opinions in positive ways, and showing sensitivity to the feelings of others."

Finally, the third type of outcome is disability status. Specifically, there are two such outcomes: whether the child is diagnosed with any disability, and whether the child is diagnosed with a cognitive disability. Cognitive disabilities include: language impairment, mental handicap, intellectual disability, and development delay. While I observe behavioral and physical disabilities in the data, I focus on cognitive disabilities because they are the most common subgroup. While for most students I observe monthly records of special-education disability status, I group monthly records into three time periods that parallel the DIBELS benchmark periods: beginning-of-year (months 1-3), middle-of-year (months 4-6), and end-of-year (months 7-9). For example, if I observe that a student is identified with a cognitive disability during any or all months 1-3, then I classify the student as having a cognitive disability during the beginning-of-year period. I operationalize each outcome as an indicator variable equal to 1 if the student is classified as having the disability, 0 otherwise.

I merge into my dataset kindergarten readiness measures that are collected either when the parent registers the student for kindergarten or when a teacher administers the WCPSS Kindergarten Initial Assessment (KIA) to the student during the first week of kindergarten. Parent-reported kindergarten readiness measures include an indicator for whether the child attended preschool for one year or more years as well as a five-point scale indicating how often the parent reads to the child (*Never*, *Once Per Month*, *Once Per Week*, *2 to 3 Times Per Week*, or *Daily*). Kindergarten readiness measures captured on the teacher-administered KIA span three domains: oral language, social-emotional skills, and physical coordination. Oral language skills include how much the student: engages in conversation, shares thoughts and ideas with others, communicates wants and needs with others, and speaks in complete sentences. Social-emotional skills include: recognizing and responding to one’s name, making appropriate choices in work and play, taking care of personal needs, effectively interacting with adults and peers, solving problems on one’s own, following school routines, responding to adult direction, and expressing feelings and show empathy to others. Finally, physical coordination consists of both gross and fine motor skills. Teachers assess motor skills based on whether the student responds appropriately when prompted to: walk forward, backward, or sideways; catch or throw a ball; and execute three-step directions with a ball. For each school readiness domain, I code the teacher’s rating as either a binary outcome or a value on a 1-4 scale. For each kindergarten cohort and within each domain, I sum the scores and standardize the total to have mean zero and unit standard deviation.

Because the outcomes vary in their availability over the timeframe of my study, I base my estimates on three different analytic samples. I refer to the first sample as the Literacy Sample, which consists of first-time kindergarteners that enrolled in the fall of the 2014-2016 school years and for whom I observe literacy outcomes over grades K-1. I refer to the second sample as the Behaviors Sample, which consists of first-time kindergarteners who enrolled in the fall of the 2015-2018 school years and for whom I observe teacher ratings of classroom behavior over grades K-1. I refer to the third sample as the Disability Sample, which consists of first-time kindergarteners who enrolled in the fall of the 2006-2016 school years and for whom I observe special-education status over grades K-3.

Table 2.1: Descriptive statistics by analytic sample and window, Wake County Public School System.

	(1)	(2)	(3)	(4)	(5)	(6)
	Literacy Sample (K-1 / 2014-17)		Behaviors Sample (K-1 / 2015-19)		Disability Sample (K-3 / 2006-19)	
Data Window	+/- 90 days	+/- 60 days	+/- 90 days	+/- 60 days	+/- 60 days	+/- 30 days
A. Sociodemographics						
Age at Cut-Off Date in Year Turned 5	5.00	5.00	5.00	5.00	5.00	5.00
Age at Cut-Off Date in Kindergarten Entry Year	5.56	5.58	5.56	5.57	5.58	5.60
Male	0.50	0.50	0.50	0.50	0.50	0.50
Black	0.18	0.18	0.19	0.18	0.19	0.19
Hispanic	0.17	0.17	0.16	0.16	0.15	0.15
Other Race	0.12	0.11	0.14	0.14	0.11	0.11
White	0.54	0.53	0.51	0.52	0.54	0.54
Ever Limited English Proficient	0.11	0.12	0.11	0.11	0.00	0.00
Classified as Special-Needs at Kindergarten Entry	-	-	-	-	0.04	0.04
B. Kindergarten Readiness Measures			-	-		
Enrolled 1+ Years in Pre-kindergarten	0.79	0.79	0.80	0.80	0.79	0.79
Parent Reads to Child						
Never	0.06	0.06	0.07	0.07	0.04	0.04
Once/Month	0.03	0.03	0.02	0.02	0.03	0.03
Once/Week	0.09	0.09	0.09	0.09	0.10	0.10
2 to 3 times/Week	0.30	0.29	0.29	0.29	0.30	0.30
Daily	0.53	0.53	0.53	0.53	0.53	0.53
Physical Readiness (Standardized)	0.10	0.10	-	-	-	-

Social-Emotional Readiness (Standardized)	0.09	0.09	-	-	-	-
Oral Language Readiness (Standardized)	0.10	0.10	-	-	-	-
C. Outcomes						
DIBELS Composite Score at First Benchmark Period (Standardized)	0.13	0.14	-	-	-	-
Ever Fails to Meet or Inconsistently Meet Conduct Expectations	-	-	0.26	0.25	-	-
Ever Fails to Meet or Inconsistently Meet Work Habits Expectations	-	-	0.34	0.33	-	-
Ever Identified with Any Disability	-	-	-	-	0.10	0.11
Ever Identified with Cognitive Disability	-	-	-	-	0.04	0.04
N Students	34278	23004	23268	15686	30774	15656

Table 2.1 reports summary statistics for each of the three analytic samples. For each sample, I report statistics based on samples restricted to students that turned five-years-old within 90, 60, or 30 days before or after the cut-off date for kindergarten entry. The size of the windows depends on the sample size. The Disability Sample, which is the largest sample, can withstand the most restrictive window of 30 days. Sociodemographic characteristics (Panel A) and kindergarten readiness measures (Panel B) are largely similar across the three samples and are stable within each sample and across windows. Means on the outcomes (Panel C) are presented but, again, are not available for every sample.

2.4 Empirical Strategy

A student's kindergarten enrollment age is potentially endogenous. Winter births are prevalent among teenagers and the unmarried (Buckles and Hungerman 2013). Redshirted kindergarteners are more likely male, White, and from affluent families (Bassok and Reardon 2013). In such situations, unobservables correlated with both kindergarten enrollment age and the outcome could bias the estimated effect of age at kindergarten entry on the outcome. Therefore, estimating the causal effect of being older at kindergarten entry on student outcomes can be difficult.

To overcome these challenges, I take the same approach as many prior studies, focusing on a policy common across many states that governs student birthdate requirements for kindergarten entry. In North Carolina, children are permitted to enroll in kindergarten in the year they turn five-years-old if they reach this age before a particular date, which I refer to as the *cut-off date*.[8] Children turning five-years-old after the cut-off date must wait until the following school year to enter kindergarten. To recover plausibly causal estimates of the effect

[8] Prior to the 2010 school year, the North Carolina cut-off date was October 16; in subsequent years, August 31.

kindergarten enrollment age on student outcomes, I employ an RD approach to exploit plausibly exogenous variation in kindergarten enrollment age created by the cut-off date. This date assigns students as-if randomly to being older kindergarten entrants (treated group) or younger kindergarten entrants (comparison group). By comparing students turning five-years-old just around the cut-off date, the RD design compares students who are equivalent in expectation on observable and unobservable characteristics except for the fact that some students turned five-years-old before or after the cut-off date. Therefore, any difference in the outcome is attributable to the difference in kindergarten enrollment age.

To exploit plausibly exogenous variation in kindergarten enrollment age resulting from the cut-off date, one can estimate parametric specifications of the following form:

$$(1)\ y_i = \alpha + \beta OlderKEntrant_i + \gamma DaysFromCut_i + \theta(OlderKEntrant_i * DaysFromCut_i) + \delta X_i + \partial R_i + \varepsilon_i.$$

y is the outcome for student *i*. *OlderKEntrant* is an indicator variable that equals 1 if in the calendar year in which student *i* enrolled in kindergarten the student was older than the oldest student that enrolled in kindergarten and turned five-years-old that same calendar year.[9] Conversely, *OlderKEntrant* equals 0 if in the calendar year in which student *i* enrolled in kindergarten the student was the same age as, or younger than, the oldest student that enrolled in kindergarten and turned five-years-old that same calendar year. *DaysFromCut* denotes the number of calendar days from the child’s fifth birthday to the cut-off date for kindergarten entry. *DaysFromCut* is centered at zero, which means negative values denote turning five-years-old before the cut-off date; positive values, turning five-years-old after the cut-off date. *DaysFromCut* enters the model linearly but is also interacted with the treatment indicator

[9] The oldest student would have turned five-years-old on January 1 of this calendar year.

OlderKEntrant to allow the relationship between the birthdate and the outcome to vary to the left and the right of the cut-off date. X denotes a vector of year fixed effects and observable student sociodemographic characteristics that include student gender, limited English proficiency (LEP), and race (Black, Hispanic, Other, White [omitted]). R denotes the vector of kindergarten readiness variables described above. ε denotes the stochastic error term. The parameter of interest in Equation (1) is β, which denotes the causal effect of being an older kindergarten entrant on the outcome, relative to being a younger kindergarten entrant.

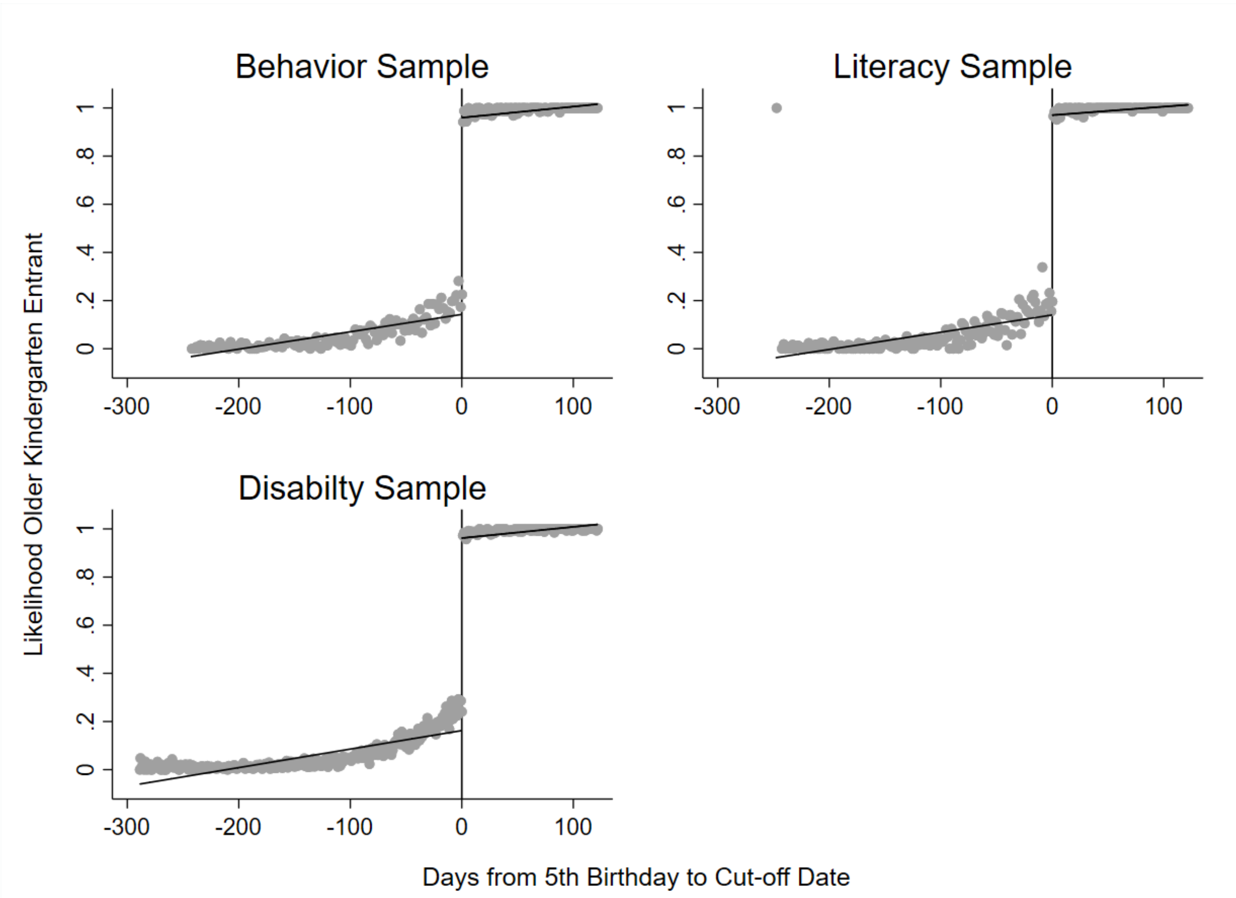
I recover and report nonparametric RD estimates using the data-driven procedure outlined in Calonico et al. (2017) that selects the optimal bandwidth around the cut-off date for kindergarten entry, which in turn minimizes the bias-variance tradeoff resulting from the use of observations closest to the cut-off versus further away.[10] I also use a triangular kernel to give greater weight to observations closest to the cut-off and less weight to observations further away. As I show in the RD plots in the Results section below, there is imperfect compliance to the left of the cut-off date for kindergarten entry and in part to the right of this date. In the year they turn five-years-old, some students turning five-years-old before the cut-off date nevertheless enroll in kindergarten during the following school year. Therefore, I implement a fuzzy RD specification. Moreover, while in principle the RD design ensures that treated students and comparison students are similar in expectation on unobservable and observable characteristics, controlling for measures of kindergarten readiness allows me to add a second layer of robustness. These measures account for parent and teacher inputs into children's skills made before the student enrolled in kindergarten. This is novel in the context of studies on kindergarten enrollment age

[10] I use the Stata package *rdrobust*.

and early elementary school outcomes, because measures of pre-kindergarten human capital inputs are not readily available in administrative data nor in survey datasets such as the ECLS-K.

2.5 Results

Figure 2.1: Regression discontinuity in probability of enrolling in kindergarten at an older age relative to peers in same kindergarten cohort.



Notes: The Literacy Sample consists of kindergarteners entering in the fall of school years 2014-16; the Behaviors Sample, kindergarteners entering in the fall of school years 2015-18; the Disability Sample, kindergarteners entering in the fall of school years 2006-16.

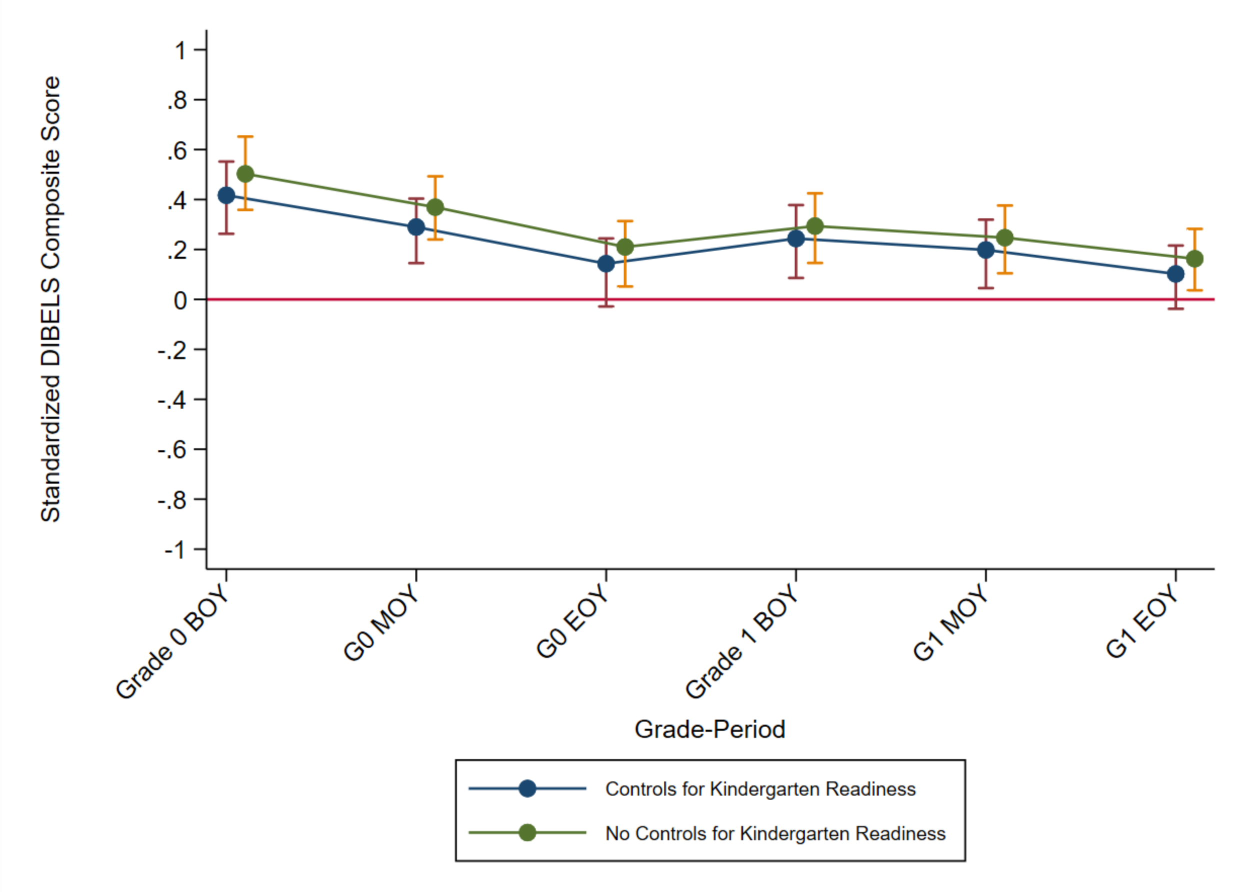
Figure 2.1 presents the extent of compliance with the kindergarten enrollment cut-off date for each of the three analytic samples. If compliance were perfect, the discontinuity in the probability of kindergarten enrollment at an older age would be unity. However, the discontinuity is between 0.70 and 0.80. For every analytic sample, the discontinuity is driven by students turning five-years-old before the cut-off date who nevertheless enter kindergarten the following year at an older age relative to their peers in the same kindergarten cohort. The majority of students turning five-years-old after the cut-off date are older than their peers who entered kindergarten in the same year in which they turned five-years-old.

Figures A3.1-A3.3 plots student-level covariates as a function of student birthdate relative to the kindergarten cut-off date. The covariates are dummy variables and include the following: ever limited English proficient in kindergarten; ever eligible for special-education services in kindergarten; male; and Black, Hispanic, other race, or White.[11] A meaningful discontinuity in a covariate at the cut-off undermines the RD approach. Such a discontinuity would raise concerns that the effect of older-age kindergarten enrollment on the outcome operates through that covariate. Fortunately, for every analytic sample, all covariates are smooth around the cut-off.

[11] WCPSS policy precludes the use of student-level free- or eligible lunch status in any analytic project.

2.5.1 Effects on K-1 Literacy

Figure 2.2: DIBELS composite score by benchmark period, grades K-1.



Notes: Each point represents the estimated effect of kindergarten enrollment at an older age on the standardized DIBELS Composite Score, recovered from the non-parametric RD specification estimated over the Literacy Sample. Controls for kindergarten readiness include: pre-kindergarten attendance; frequency parent reads to child; and measures of oral language skills, social-emotional skills, and physical coordination. Additional control variables include for gender, race, limited English proficiency, and school year fixed effects. Grade 0 refers to kindergarten. BOY, MOY, and EOY denote beginning-, middle-, and end-of-year, respectively. Robust 95% confidence intervals.

Figure 2.2 presents the effect of kindergarten enrollment at an older age on the DIBELS composite score at each benchmark period during grades K-1. (Coefficient estimates and robust standard errors are presented in Table B.1.) Older kindergarten entrants maintain an advantage in foundational literacy skills over their younger counterparts at the beginning- and middle-of-year benchmark periods, in both kindergarten and first grade. However, this advantage diminishes rather quickly in each grade—at a faster rate during kindergarten and a slightly slower rate during first grade. Notably, the advantage in foundational literacy skills had by older kindergarten entrants is indistinguishable from zero at the end of the year in each grade, when the estimate is conditioned on the rich set of kindergarten readiness measures. In other words, after accounting for human capital investments made prior to children entering kindergarten, older and younger kindergarten entrants are similar in foundational literacy skills by the end of kindergarten and again by first grade. This finding, which is consistent with prior studies reporting that after first grade test scores converge between older and younger kindergarten entrants (Lubotsky and Kaestner 2016), suggests that the age advantage in literacy had by older entrants fades out even earlier than once thought. To be sure, Figure 2.2 does not indicate whether younger kindergarten entrants grow in their foundational literacy skills or older kindergarten entrants decline in their skills over these two benchmark periods.

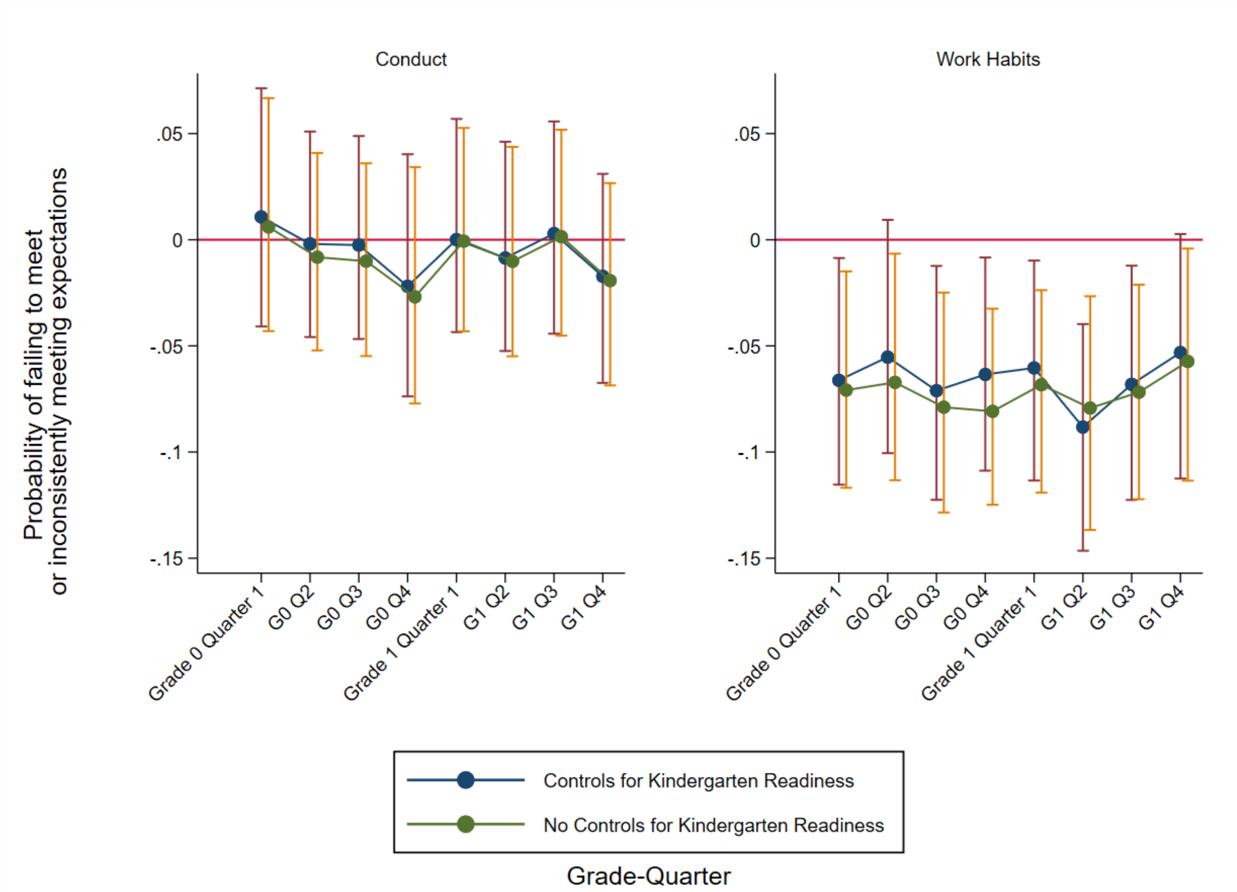
An additional finding is the widening of the advantage had by older kindergarten entrants in foundational literacy skills during the summer break, that is, between kindergarten end-of-year and first grade beginning-of-year. Over the summer, perhaps older kindergarten entrants are more robust to summer learning loss—and better preserve their foundational literacy skills—compared to their younger counterparts. One important feature of the DIBELS as a screener is that three out of the four the subtests administered at the end of kindergarten are also administered at the start of first grade. The continuity in subtests between these two benchmark periods should assuage concerns that the change in the advantage in foundational literacy skills had by older kindergarten entrants is merely an artifact of the DIBELS screener.

Some have raised concerns about the predictive validity of the DIBELS screener (e.g., Goodman 2006, Manzo 2005, Strauss 2007). Within the context of the current study, predictive validity is the correlation between a student's performance on the DIBELS and the student's performance on a criterion measure, or a "gold standard assessment," such as a summative, state criterion-referenced high-stakes test (Clemens et al., 2019). At the heart of the issue is how meaningful are DIBELS composite scores, at least in terms of how much they predict future reading achievement. To assuage these concerns, I examine how much DIBELS composite scores predict scores on the North Carolina third-grade end-of-grade summative reading assessment (EOG3R).

Table B.1 presents the results from an OLS regression model in which the outcome is the EOG3R score and the predictor variable of interest is the DIBELS composite score at a specific grade-benchmark period. The unconditional estimates are large, highly significant at conventional levels, and measured with considerable precision (Column 1). These estimates shrink in magnitude somewhat upon conditioning the model on student sociodemographic characteristics and measures of kindergarten readiness (Column 2). Intuitively, the estimates grow the closer in time is the DIBELS composite score to the EOG3R score. Overall, these results suggest that the DIBELS screener is a credible measure of foundational literacy skills that correlates meaningfully to later measures of literacy and reading achievement.

2.5.2 Effects on K-1 Classroom Behaviors

Figure 2.3: Quarterly Conduct and Work Habits teacher ratings, grades K-1.



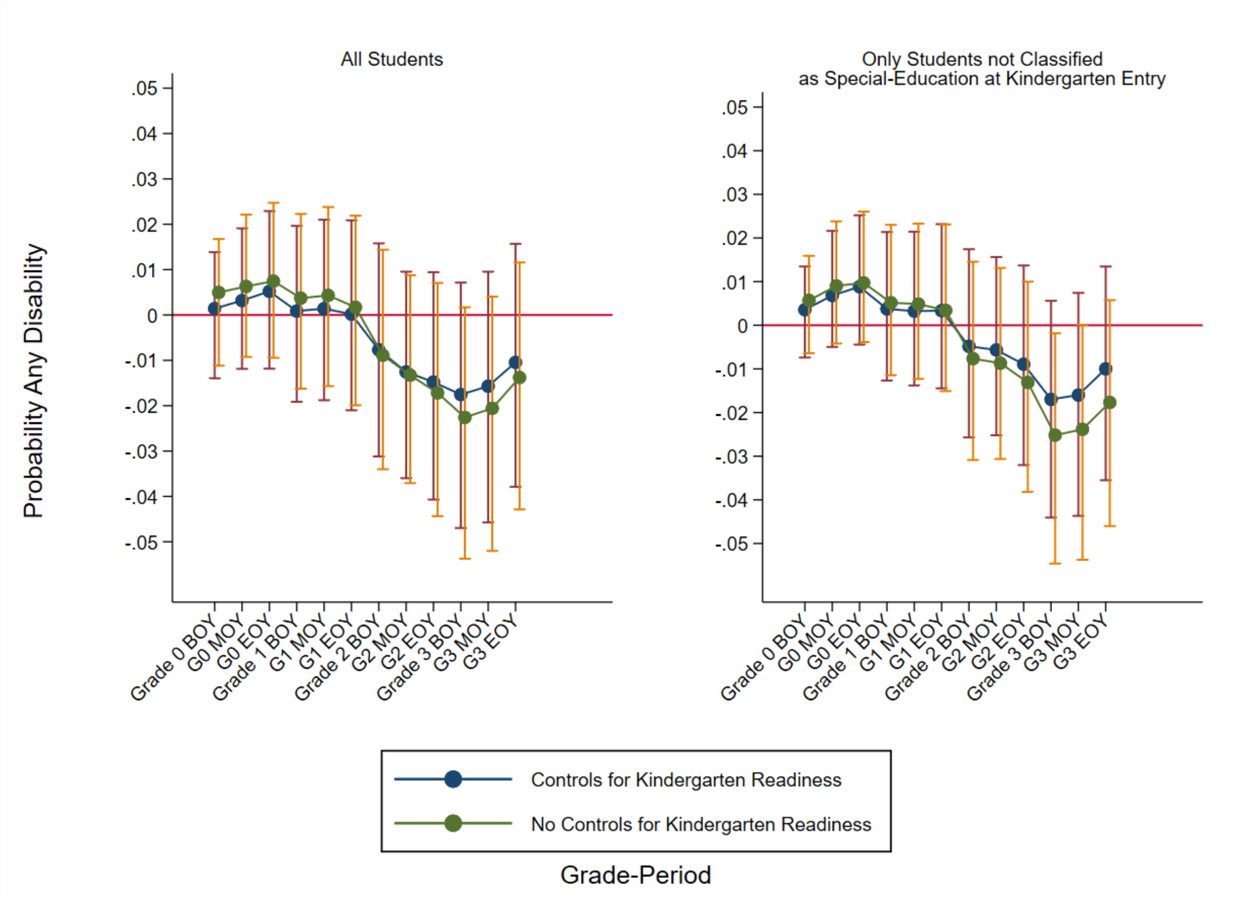
Notes: Conduct refers to how much the student cooperates with others, respects others, and observes rules and procedures. Work Habits refers to how much the student uses time wisely, listens carefully, completes assignments, writes legibly, works independently or seeks help when needed, and completes work. Each point represents the estimated effect of kindergarten enrollment at an older age on the Conduct or Work Habits indicator, recovered from the non-parametric RD specification estimated over the Behaviors Sample. Controls for kindergarten readiness include pre-kindergarten attendance and the frequency parent reads to child. Additional control variables include: gender; race; limited English proficiency; special-education status; and school year fixed effects. Grade 0 refers to kindergarten; there are 4 quarters per grade. Robust 95% confidence intervals.

Figure 2.3 presents the effect of kindergarten enrollment at an older age on teacher ratings of how often students inconsistently meet or fail to meet expectations for the Conduct and Work Habits behavioral measures. (Coefficient estimates and robust standard errors are presented in Table B.2.) On the one hand, the left panel indicates that during grades K-1 teachers perceive no difference in the Conduct measure between older and younger kindergarten entrants. This pattern persists throughout these two grade levels. On the other hand, the right panel indicates that teachers perceive older kindergarten entrants as slightly more skilled on the Work Habits measure than younger kindergarten entrants. In the first quarter of kindergarten, for example, older kindergarten entrants are about 6 percentage points less likely rated by teachers as failing to meet or inconsistently meeting the Work Habits expectation. The advantage had by older kindergarten entrants in Work Habits persists throughout both kindergarten and first grade and does not change meaningfully.

It is notable that the Conduct and Work Habits ratings exhibit patterns inconsistent with one another. This inconsistency can be explained, at least in part, by the fact that Conduct and Work Habits capture behaviors that students, in virtue of their age, are likely to have honed to different degrees at each grade-quarter combination. Conduct is a relational measure insofar as a student's Conduct rating is based on how well the student cooperates with the student's peers. Work Habits is an "individual" measure insofar as it is based largely on how well the student functions independently in the classroom vis-à-vis completing tasks alone. Another explanation for this result is that both younger and older kindergarten entrants enter school with more or less the same level of understanding of how to work with their peers in a kindergarten classroom setting, a setting likely more structured than at-home or pre-kindergarten settings. In contrast, older kindergarten entrants may enter this grade more adept than their younger counterparts at completing tasks independently, especially if they have spent extra time in pre-kindergarten settings. Data limitations preclude me from controlling for kindergarten readiness in the physical and social-emotional domains, which would likely explain more of the variance in Work Habits ratings.

2.5.3 Effects on K-3 Disability Status

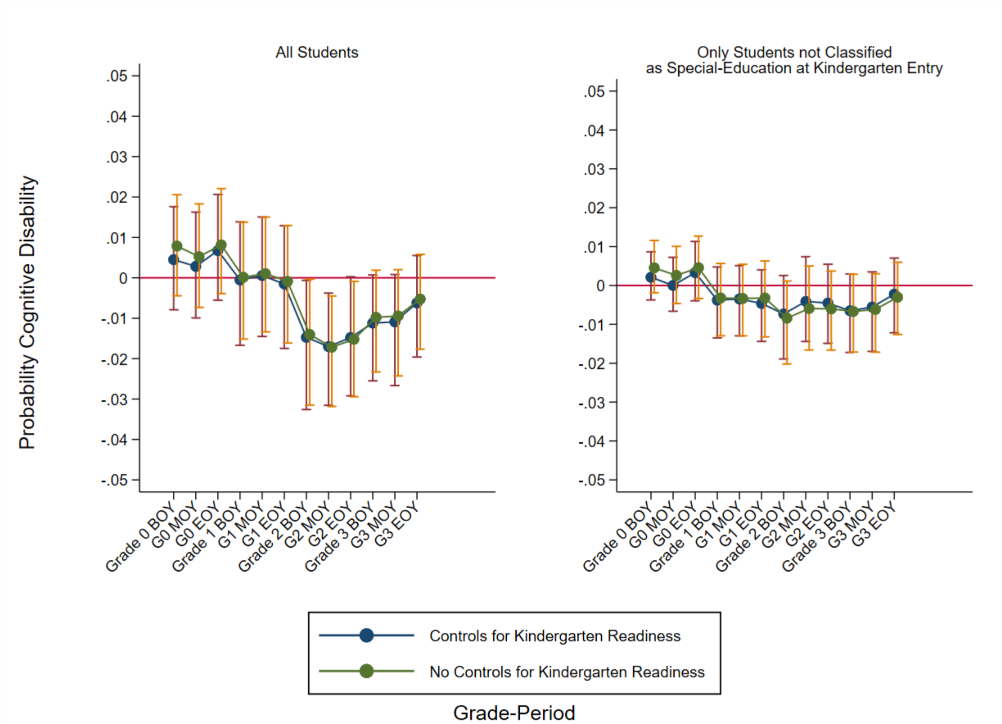
Figure 2.4: Probability of any disability identification, by benchmark period, grades K-3.



Notes: Each point represents the estimated effect of kindergarten enrollment at an older age on the likelihood of being identified with any disability, recovered from the non-parametric RD specification estimated over the Disability Sample. Controls for kindergarten readiness include pre-kindergarten attendance and frequency parent reads to child. Additional control variables include: gender; race; limited English proficiency; and school year fixed effects. Grade 0 refers to kindergarten. BOY, MOY, and EOY denote beginning- (months 1-3), middle- (months 4-6), and end-of-year (months 7-9), respectively. Robust 95% confidence intervals.

Figure 2.4 presents the effect of kindergarten enrollment at an older age on whether the student is identified as having any disability (e.g., physical, emotional, cognitive). (Coefficient estimates and robust standard errors are presented in Table B.3.) The confidence intervals in the left panel include zero, which implies that older kindergarten entrants are as likely as their younger counterparts to be identified as having a disability during grades K-3. Reassuringly, the magnitude of these estimates is consistent with those reported in recent studies documenting a 3-5 percentage point lower likelihood of disability identification during 3rd-8th grade for older kindergarten entrants (Dhuey et al. 2019). To ensure that my estimates are based on students who were not already identified as requiring special education services at kindergarten entry—and who therefore would likely be identified with a disability in later grades—the right panel excludes such students. The right panel presents coefficient estimates that are qualitatively similar to those in the left panel, although they are slightly attenuated. This finding is notable because it suggests that older kindergarten entrants who have no known prior disability are as likely as younger kindergarten entrants to be identified as having a disability during grades K-3.

Figure 2.5: Probability of cognitive disability identification, by benchmark period, grades K-3.



Notes: Each point represents the estimated effect of kindergarten enrollment at an older age on the likelihood of being identified with a cognitive disability (i.e., language impairment, mental handicap, intellectual disability, and development delay), recovered from the non-parametric RD specification estimated over the Disability Sample. Controls for kindergarten readiness include pre-kindergarten attendance and frequency parent reads to child. Additional control variables include: gender; race; limited English proficiency; and school year fixed effects. Grade 0 refers to kindergarten. BOY, MOY, and EOY denote beginning- (months 1-3), middle- (months 4-6), and end-of-year (months 7-9), respectively. Robust 95% confidence intervals.

Figure 2.5 extends the preceding analysis by presenting the effect of kindergarten enrollment at an older age on whether the student is identified as having a cognitive disability. The left panel shows that older kindergarten entrants are as likely as their younger counterparts to be identified as having a cognitive disability during kindergarten and first grade. However, at the beginning of second grade, older kindergarten entrants are significantly less likely than their younger counterparts to be identified as having a cognitive disability, although this advantage fades away by third grade. The right panel shows that, upon excluding kindergarten entrants who were identified with special-needs at the start of kindergarten, older entrants are as likely as younger entrants to be identified as having a cognitive disability during grades K-3.

I also consider the effect of older-age kindergarten enrollment on identification with two specific types of disability: learning disability (Figure A3.4) and speech impairment (Figure A3.5). In large part, the patterns in these figures resemble those in the figures that plotted any disability identification and cognitive disability identification in relation to the grade-period. I note, however, that the likelihood of learning disability identification among older kindergarten entrants relative to their younger counterparts is significantly larger throughout third grade.

2.6 Discussion and Conclusion

I examined how much older-age kindergarten enrollment affects foundational literacy skills, teacher ratings of student behaviors, and disability identification, all within and across grades K-3. Conditional on measures of kindergarten readiness, older kindergarten entrants are ahead of their younger counterparts in terms of foundational literacy skills at every K-1 benchmark period except at end-of-year. Over these same grades, older kindergarten entrants also consistently meet Work Habits expectations more often than their younger counterparts, although both groups are as likely as one another to meet Conduct expectations. Finally, older and younger kindergarten entrants face similar rates of disability identification during grades K-3. Importantly, these findings are conditional on measures of kindergarten readiness.

I speculate about why some of the results that I presented above appear inconsistent with those documented in prior studies. First, whereas Lubotsky and Kaestner (2016) reported that older and younger kindergarten entrants start to resemble one another on measures of reading achievement after first grade, I found that these students resemble one another in foundational literacy skills as early as the end of kindergarten. These results are not inconsistent. Rather, the DIBELS detects skills that are different from those measured in widely available measures of reading achievement (e.g., the Peabody Individual Achievement Test administered as part of the ECLS-K, end-of-grade standardized reading achievement tests), although to be sure these skills overlap somewhat (see Table B.1 for how well the DIBELS composite score predicts 3rd grade reading achievement). Because the returns to human capital investments decline as students age, and because skill formation enhances future skill formation, I contend that this particular finding is relevant to policymakers, especially those considering delayed kindergarten enrollment as an approach to closing academic achievement gaps that emerge early and grow among select student subgroups (e.g., low-income) or shoring up skill formation for student groups susceptible to learning loss.

Second, whereas Lubotsky and Kaestner (2016) reported that older kindergarten entrants are similar to their younger counterparts on a wide set of noncognitive outcomes at the start of kindergarten, I find that across grades K-1 older entrants are advantaged in terms of their teachers' perceptions of their ability to work independently. That the instruments measure different skills is an unsatisfying answer. In fact, the kinds of skill captured in WCPSS's Work Habits score appear, at least in part, on the ECLS-K, specifically in the items from the

Approaches to Learning measure. One possibility I cannot rule out in the current study is that teachers are aware of student relative age—that is, whether a student entered kindergarten at an older age—and that teachers are biased toward viewing older kindergarten entrants as more adept at engaging in independent work. This possibility is probably slim. During grades K-1, older kindergarten entrants are nearly always rated more favorably than their younger counterparts in their ability to work independently.

Third, whereas collectively prior studies document a lower likelihood of disability identification among older kindergarten entrants relative to their younger counterparts (Dhuey et al. 2019; Dhuey and Lipscomb 2010; Shapiro 2020) over grades K-8, I document little to no advantage in the WCPSS data. Most convincing are the analyses that exclude students who are identified at kindergarten entry as requiring special education services. By excluding students who would have very well been identified with a disability in subsequent grades, these analyses offer an uncontaminated estimate of the desired treatment effect. One avenue for future research could explore cross-context differences in disability identification policies, including processes for the identification of certain sub-category disabilities (e.g., speech impairment).

The current study begins to fill gaps in the literature around within-grade differences in cognitive and noncognitive outcomes between older and younger kindergarten entrants. One objective for future research could be to implement the analyses in this study on data from other states or local contexts, even if some of the measures are not identical to those in this study. Different measures might even be preferrable, as they will contribute to a holistic understanding of the effects of kindergarten enrollment age on early elementary school outcomes. Given the widespread implementation of SLDS, this objective appears quite reasonable.

REFERENCES

Bassok, Daphna, and Sean F. Reardon. 2013. "'Academic Redshirting' in Kindergarten: Prevalence, Patterns, and Implications." *Educational Evaluation and Policy Analysis* 35 (3): 283–97. http://www.jstor.org/stable/43773434.

Buckles, Kasey S., and Daniel M. Hungerman. 2013. "Season of Birth and Later Outcomes: Old Questions, New Answers." *The Review of Economics and Statistics* 95 (3): 711–24. https://doi.org/10.1162/REST_a_00314.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik. 2017. "Rdrobust: Software for Regression-Discontinuity Designs." *The Stata Journal: Promoting Communications on Statistics and Stata* 17 (2): 372–404. https://doi.org/10.1177/1536867X1701700208.

Clemens, Nathan H., Kelsey Ragan, and Christopher Pricket. 2018. "Predictive Validity." In *The Sage Encyclopedia of Educational Research, Measurement, and Evaluation*, edited by Bruce B. Frey. Los Angeles: SAGE Reference.

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31–47. https://doi.org/10.1257/aer.97.2.31.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78 (3): 883–931. https://doi.org/10.3982/ECTA6551.

Datar, Ashlesha. 2006. "Does Delaying Kindergarten Entrance Give Children a Head Start?" *Economics of Education Review* 25 (1): 43–62. https://doi.org/10.1016/j.econedurev.2004.10.004.

Deming, David, and Susan Dynarski. 2008. "The Lengthening of Childhood." *Journal of Economic Perspectives* 22 (3): 71–92.

Dhuey, Elizabeth, David Figlio, Krzysztof Karbownik, and Jeffrey Roth. 2019. "School Starting Age and Cognitive Development." *Journal of Policy Analysis and Management* 38 (3): 538–78. https://doi.org/10.1002/pam.22135.

Dhuey, Elizabeth, and Stephen Lipscomb. 2010. "Disabled or Young? Relative Age and Special Education Diagnoses in Schools." *Economics of Education Review* 29 (5): 857–72. https://doi.org/10.1016/j.econedurev.2010.03.006.

Elder, Todd E., and Darren H. Lubotsky. 2009. "Kindergarten Entrance Age and Children's Achievement: Impacts of State Policies, Family Background, and Peers." *The Journal of Human Resources* 44 (3): 641–83. http://www.jstor.org/stable/20648912.

Fuchs, Lynn S., Douglas Fuchs, Donald L. Compton, Joan D. Bryant, Carol L. Hamlett, and Pamela M. Seethaler. 2007. "Mathematics Screening and Progress Monitoring at First Grade: Implications for Responsiveness to Intervention." *Exceptional Children* 73 (3): 311–30. https://doi.org/10.1177/001440290707300303.

Goodman, Kenneth S., ed. 2006. *The Truth about DIBELS: What It Is, What It Does*. Portsmouth, NH: Heinemann.

Hemelt, Steven W., and Rachel B. Rosen. 2016. "School Entry, Compulsory Schooling, and Human Capital Accumulation: Evidence from Michigan." *The B.E. Journal of Economic Analysis & Policy* 16 (4). https://doi.org/10.1515/bejeap-2015-0219.

Hill, Carolyn J., Howard S. Bloom, Alison Rebeck Black, and Mark W. Lipsey. 2008. "Empirical Benchmarks for Interpreting Effect Sizes in Research." *Child Development Perspectives* 2 (3): 172–77. https://doi.org/10.1111/j.1750-8606.2008.00061.x.

Jenkins, Joseph R., Roxanne F. Hudson, and Evelyn S. Johnson. 2007. "Screening for At-Risk Readers in a Response to Intervention Framework." *School Psychology Review* 36 (4): 582–600.

Lubotsky, Darren, and Robert Kaestner. 2016. "Do `Skills Beget Skills'? Evidence on the Effect of Kindergarten Entrance Age on the Evolution of Cognitive and Non-Cognitive Skill Gaps in Childhood." *Economics of Education Review* 53 (August): 194–206. https://doi.org/10.1016/j.econedurev.2016.04.001.

Manzo, Kathleen Kennedy. 2005. "National Clout of DIBELS Test Draws Scrutiny." *Education Week*, 2005.

Shapiro, Anna. 2020. "Over Diagnosed or Over Looked? The Effect of Age at Time of School Entry On Students Receiving Special Education Services." EdWorkingPapers.com. https://www.edworkingpapers.com/ai20-259.

Strauss, Valerie. 2007. "DIBELS Test: A Question Of Validity," March 26, 2007. http://www.washingtonpost.com/wp-dyn/content/article/2007/03/25/AR2007032501082.html.

Tourangeau, Karen, Christine Nord, Thanh Lê, Alberto G. Sorongon, and Michelle Najarian. 2009. "Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K): Combined User's Manual for the ECLS-K Eighth-Grade and K-8 Full Sample Data Files and Electronic Codebooks. NCES 2009-004." Washington, D.C: National Center for Education Statistics. https://files.eric.ed.gov/fulltext/ED511826.pdf.

CHAPTER 3: TEACHER ABSENCES, TRANSFERS, AND TURNOVER UNDER SCHOOL DESEGREGATION: EVIDENCE FROM A CONTEMPORARY STUDENT REASSIGNMENT PLAN

3.1 Overview

School desegregation plans diversify schools on a variety of student characteristics. Teachers in schools that implement these plans might exit their schools or seek positions in other schools where student sociodemographic characteristics align with their preferences for where to work. If the diversification of student skills poses challenges to the delivery of instruction, then these plans could also lower teacher productivity. To investigate these possibilities, we construct a unique teacher-level panel dataset that includes applications to transfer and records of absences for teachers in the Wake County Public School System, which implemented a novel student reassignment plan during the 2000s that integrated schools socioeconomic disadvantage and achievement. Using school and year fixed effects models, we recover a precisely estimated null effect of the reassignment plan on the likelihood that a teacher resigns or seeks to transfer, as well as a modest but significant drop in teacher sick days. Our findings expand the body of evidence on contemporary school desegregation plans to include teacher outcomes, which may be helpful for policymakers considering implementing these plans in their own contexts.

3.2 Introduction

In recent years, school districts have implemented policies similar in spirit to those of court-ordered desegregation plans realized in the wake of *Brown v. Board of Education*. The Berkeley Unified School District redrew attendance boundaries so that students living in majority White and majority non-White residential areas attend the same schools. Public schools in certain New York City neighborhoods grant priority to English Language Learners (ELLs) in the admissions process to gifted and talented elementary school programs. One aim of these policies is to improve student outcomes, especially achievement for students from less-advantaged backgrounds who may otherwise attend historically underperforming schools that serve student populations that are largely homogeneous in terms of race and family income. Yet little is known about how these policies affect teachers, a critical component of the education production function and through which school desegregation might influence student outcomes.

Studies on the relationship between student reassignment and student outcomes motivate our study, in which we focus on whether, and how much, school desegregation affects teacher outcomes. Collectively, these studies suggest that student reassignment is disruptive to instruction and learning, although the findings are mixed. On the one hand, studies directly examine the effects of school desegregation on student outcomes. Nationwide, these orders were associated with lower Black dropout rates and little or no effect on White dropout rates (Guryan 2004; Johnson 2011). In Boston, achievement among receiving-school students at relatively advantaged elementary schools was largely unaffected by their new peers from less-advantaged schools (Angrist and Lang 2004). However, following the end of race-based busing in Charlotte, White and minority students scored lower on math tests when assigned to schools with relatively large shares of minority students (Billings, Deming, and Rockoff 2014). This pattern complements findings from Texas, where larger shares of Black students were associated with lower math achievement for Blacks (Hanushek, Kain, and Rivkin 2009). On the other hand, studies investigate school closures and their effects on student outcomes. School closures appear to have negative spillovers on achievement for receiving-school students that persist for at least two years after closure (Brummet 2014; Steinberg and MacDonald 2019), although other studies report no such effects (Engberg et al. 2012) or a negative effect lasting only one year after closure (Gordon et al. 2018). Absences among receiving-school students also increase in the share of displaced students (Steinberg and MacDonald 2019), although other studies report no such effect (Engberg et al. 2012). On the whole, these results suggest that student reassignment policies may have negative effects on student outcomes that the policies are designed to improve.

We argue that one explanation for the negative effects summarized above is that the assignment of students to schools that they had not previously attended could abruptly alter the school-level sociodemographic characteristics of receiving-schools, thereby broadening the characteristics of the average student to which teachers must provide instruction. These student characteristics almost certainly track others, some of which are unobservable to the researcher (e.g., family human capital investments, student motivation for schoolwork). Importantly, some of these student characteristics bear on the quantity and quality of instruction and learning. Teachers in receiving-schools who must instruct a more skill-diverse student population relative to prior years may face new challenges on several job dimensions, such as classroom management and the delivery of effective instruction. In response, these teachers might take actions that have negative effects on student outcomes and that ultimately undermine the policy goals of student reassignment plans that aim to desegregate schools.

To investigate the relationship between student reassignment and teacher outcomes, we study whether, and how much, school desegregation influences teachers' decisions about where they work and how productive they are. We study an innovative socioeconomic desegregation plan implemented during the 2000s in the Wake County Public School System (WCPSS),[12]

[12] The full text of the assignment policy is available at: https://boardpolicyonline.com/bl/?b=wake_old#&&hs=189696. Parcel and Taylor (2015) also provide a nice review of the WCPSS reassignment plan.

historically the largest school district in North Carolina in terms of student enrollment. The central feature of the Wake Reassignment Plan[13] was the wholesale assignment of students--based on a kind of neighborhood unit--to base schools that were different from the schools these students had been assigned in the previous year. We treat a between-year change in base school assignment as *reassignment*. The purpose of the Reassignment Plan was to balance the distribution of students across schools in terms of socioeconomic disadvantage and achievement. We describe the Plan in greater detail below. Importantly, the Reassignment Plan resulted in plausibly exogenous shocks to school-level sociodemographics, which we argue had the potential to influence three teacher outcomes, specifically turnover, applier status (i.e., whether the teacher applied to transfer), and absences. We address four research questions:

1. Relative to years when a school was unaffected by the student assignment procedures of the Wake Reassignment Plan (*non-reassignment years*), did teachers in years when students were involuntarily assigned to the school under the Plan (*reassignment years*):
    a. Resign at higher rates?
    b. Apply to transfer to other schools at higher rates?
    c. Take more sick days?
2. Did the effects on teacher outcomes vary by measures of teacher quality, namely National Board for Professional Teaching Standards certification and years of experience?

A large body of literature on teacher mobility strongly suggests that teachers generally prefer to work in schools that serve affluent, White, and high-achieving students (Boyd et al. 2005, 2011; Clotfelter, Ladd, and Vigdor 2005, 2011; Hanushek et al. 2004; Ingersoll and May

[13] Throughout the paper we refer to the Wake Reassignment Plan, the Reassignment Plan, the Wake Plan, and simply the Plan.

2012; Scafidi et al. 2007; Steele et al. 2015). Qualitative research also describes teachers who believe that school integration leads to disciplinary problems and has little positive effect overall on receiving-school students (Caldas, Bankston, and Cain 2007). Therefore, when their own schools desegregate, teachers might apply to transfer to schools where overall student sociodemographic characteristics align more closely with their preferences. Teachers might also simply exit their schools.

Teachers might also take more days off than they had prior to the desegregation of their own schools. One reason might be to cope with new challenges associated with instructing a student population that is more skill-diverse than they had previously instructed. Importantly, teacher absences have negative effects on student achievement (Clotfelter, Ladd, and Vigdor 2009; Herrmann and Rockoff 2012; Miller, Murnane, and Willett 2007; Tingle et al. 2012) as well as direct costs to school districts estimated as high as $1,800 per year per teacher (Joseph, Waymack, and Zielaski 2014).

To preview our results, we find that, in reassignment years, teachers at a school were no more likely to resign or apply to transfer to other schools than in non-reassignment years. However, we find that teachers took fewer sick days during reassignment years relative to non-reassignment years. This pattern is potentially attributable to increases in teacher productivity in response to the challenges associated with delivering instruction to a more skill-diverse student population than that of prior years. We begin by grounding our study in the relevant literature and then describing the policy context of transfer applications and teacher absences. We then describe the data and the empirical strategy we use to estimate the effects of student reassignment under the Wake Plan on teacher resignation, applier status, and sick days. Finally, we present the findings and conclude.

3.3 Literature Review

Our study contributes to three strands of literature. First, we build on studies examining contemporary school desegregation efforts. We are aware of no other study that explores the effects of such a plan implemented in a large school district on teacher outcomes *under exposure to the plan*. Most similar to our study is Jackson (2009), who examined the effects of the elimination in 2000 of school desegregation in Charlotte-Mecklenburg Public Schools (CMS) on the distribution of teacher quality across the district in the year that followed. The undoing of the Charlotte Plan led to immediate and sizable changes in school-level student racial composition, achievement, and economic disadvantage throughout CMS. Jackson (2009) documented accompanying changes in the distribution of teacher quality across the district immediately after these changes. High-quality teachers, which Jackson (2009) distinguished based on measures such as value-added scores and *U.S. News and World Report University Rankings* of undergraduate institutions, exited schools in mostly Black neighborhoods that had once served a racial mix of students under the Charlotte Plan but then increased in Black enrollment shares after the Plan's elimination. Jackson (2009)'s study is immensely valuable to our collective understanding of teacher mobility following the elimination of a long-standing school desegregation plan.

For two reasons we suspect that applier status and resignation may have differed following the end of the Charlotte Plan and the implementation of the Wake Plan. The first is a difference in scale. The wholesale elimination of the Charlotte Plan had system-wide implications for school-level sociodemographic characteristics across one of the largest school districts in the United States. In contrast, the implementation of the Wake Plan affected only a subset of the district's schools every year. In this regard, the Wake Plan resembles contemporary school desegregation efforts in cities like Berkeley and NYC. The second reason is that the elimination of the Charlotte Plan was associated with sizable changes in sociodemographic characteristics within schools across CMS. In contrast, the Wake Plan involved the piecemeal reassignment of "neighborhood" units in certain county areas, resulting in relatively modest impacts on within-school, sociodemographic characteristics. Given these differences between the elimination of the Charlotte Plan and the implementation of the Wake Plan, teachers likely responded differently in the two contexts.

Second, our study builds on others that have examined potential teacher mobility using teachers' transfer applications rather than job histories (Barbieri, Rossetti, and Sestito 2011; Boyd et al. 2011; Engel, Jacob, and Curran 2014). Whereas a transfer application documents a teacher's interest in other schools, a teacher's job history documents where the teacher has worked, and not necessarily where the teacher would have preferred to work. As such, transfer applications document teachers' preferences for school characteristics absent contamination from demand-side factors that almost certainly influence teachers' labor supply decisions, such as job vacancies and principal recruitment of teachers to certain schools. We extend this small body of literature by exploring whether, and the extent to which, plausibly exogenous changes in school-level sociodemographic characteristics resulting from Wake Plan-induced student reassignment affect a teacher's applier status and resignation likelihood.

Our use of teachers' transfer applications renders it most similar to Boyd et al. (2011), who draw upon transfer applications for nearly 81,000 New York City (NYC) public school teachers over the 2006-07 and 2007-08 school years. Our study complements Boyd et al. (2011) in two ways. First, we study teacher outcomes in a district that may more closely resemble districts in other parts of the United States. Whereas NYC serves the largest and most racially diverse student population in the United States, WCPSS is comparable to large and geographically diffuse school districts in major--but considerably smaller--metropolitan areas that have urban centers and rural fringes. Second, we draw upon data from a long timeframe that includes years with weak and typical teacher labor markets. In contrast, Boyd et al. (2011) use data from two school years, 2007 and 2008,[14] one of which coincided with the Great Recession. During this time period, public education saw deeper budget cuts and more reductions in force than in prior years (Goldhaber et al., 2016; Knight and Strunk, 2016), which may have influenced teachers' transfer-seeking behaviors and mobility decisions more than in otherwise stable job markets. This may have been especially true for novice teachers, who are more likely than more-experienced teachers to exit the profession voluntarily or involuntarily (Fulbeck 2014; Goldhaber et al. 2016; Goldhaber, Gross, and Player 2011; Johnson, Kraft, and Papay 2012; Loeb, Darling-Hammond, and Luczack 2005; Smith and Ingersoll 2004). Budget cuts and reductions in force during the Great Recession also disrupted studies on teacher retention conducted around the United States (e.g., Wechsler et al. 2011).

Finally, our study contributes to the literature on teacher absences. These studies document a negative effect of teacher absences on student achievement (Ahn 2013; Clotfelter, Ladd, and Vigdor 2009; Herrmann and Rockoff 2012; Miller, Murnane, and Willett 2007; Tingle et al. 2012); demonstrate that federal- and school-level accountability policies lower the number of days off that teachers consume (Gershenson 2016; Jacob 2013); and show that workplace conditions (Ost and Schiman 2017) and teachers' own sociodemographic characteristics (Rosenblatt and Shirom 2005) predict teacher absences. We extend these studies by considering

[14] We refer to the school year using the calendar year of the spring term (e.g., 2007 refers to the 2006-07 school year).

whether teacher absences increase in response to changes in overall student characteristics due to student reassignment under the Wake Reassignment Plan.

We suspect that changes in school-level sociodemographic characteristics induced by the Wake Reassignment Plan had the potential to raise workplace demands and stressors, which in turn would have raised teacher absences. In other industries, workers that experience stressful workplaces demonstrate higher rates of mental and physiological illnesses (Nixon et al. 2011; Sonnentag and Frese 2003) and take more days off relative to their peers in less stressful workplaces (Darr and Johns 2008). More directly, a large literature on teacher stress and burnout also supports our line of inquiry. Teaching is one of the most stressful professions (Chang 2009; Schonert-Reichl 2017), a fact that has spurred interest in interventions to help teachers cope with workplace stressors (e.g., Naghieh et al. 2015; Roeser et al. 2013). We treat changes in school-level sociodemographic characteristics induced by the Wake Plan as having had the potential to intensify workplace demands and teacher workload through the diversification of student skill levels. As a result, teachers in receiving-schools may have used absences to seek respite from challenging teaching assignments and classroom environments.

3.4 Background

3.4.1 WCPSS and School Integration through Student Assignment[15]

WCPSS is divided into roughly 1,000 geographic administrative units for the purposes of student assignment. Each unit, which we refer to as a *node*, contains about 125 students from different grade levels. We refer to the node-grade combination as a *node-grade unit*. During 2001-10, the Wake Reassignment Plan outlined two criteria that governed annual student

[15] Others document the historical and political origins of the Wake Reassignment Plan (see Parcel and Taylor 2015) as well as the mechanics of the Plan (see Ayscue et al. 2018; Carlson et al. 2020; Williams and Houck 2013; Wake Education Partnership 2003). We restrict our discussion to the facts most relevant to our study.

assignment to schools: (a) no more than 40 percent of students assigned to any school can be eligible for free or reduced-price lunch (FRL); and (b) no more than 25 percent of students assigned to any school can perform below grade level, measured by end-of-year state accountability tests. Each year, the WCPSS Office of Student Assignment (OSA) met these criteria in two ways. The first was through business-as-usual assignment practices, whereby students in a node-grade unit were assigned to the same base elementary, middle, or high school to which they were assigned the year before. The second was through the assignment of a number of node-grade units--and thus all of the students residing therein--to schools that differed from those to which the students in these node-grade units were assigned the year before. We refer to these schools as *alternate base schools*. During the Plan years, the OSA assigned roughly 25 percent of node-grade units to alternate base schools at least once. In principle, the OSA would consider a base school's expected sociodemographic characteristics under business-as-usual student assignment practices and reallocate students as needed to meet the two criteria above.[16]

Therefore, every year during which WCPSS implemented the Reassignment Plan, the OSA effectively randomly assigned the *reassignment status* of a school--and the teachers therein--to one of two conditions. In the first condition, the OSA assigned node-grade units to the school according to ordinary, business-as-usual student assignment practices. Under this first condition, which we think of as the comparison condition, the school did not serve as an alternate base school. In the second condition, the OSA assigned node-grade units to alternate base schools under the policy goals of the Plan, that is, to satisfy the two criteria outlined in the Wake

[16] The Wake Reassignment Plan also listed listed 5 additional factors that would be considered in the school assignment process (Wake Education Partnership, 2003): (1) instructional program; (2) consistency with elementary, middle, and high school grade ranges; (3) facility capacity; (4) stability for families; and (5) proximity.

Reassignment Plan. We think of this second condition as the *treatment condition*. If treated, the base school became an alternate base school because the school was assigned students from node-grade units that had not been assigned to the school the year before.

Typically, the Wake Reassignment Plan reallocated students from less affluent, urban schools to affluent, suburban schools. Thus, the Plan allowed the district to strive for an equitable distribution of economically disadvantaged and low-achieving students across the district. This is evident if one compares the Dissimilarity Index for CMS against that of WCPSS. The Dissimilarity Index is a continuous measure of school segregation in a defined area (e.g., a school district) with the values 0-1 to represent the share of students who would need to be reallocated to ensure an even distribution of those students across schools (see Massey and Denton 1988). Under the Wake Plan, WCPSS's Black-White Dissimilarity Index fluctuated between 0.25 and 0.33, meaning in any given year WCPSS would have had to reallocate 33 percent of Black students to achieve an even distribution of these students across the district (Williams and Houck 2013). In contrast, CMS, which eliminated its race-based desegregation policy in 2002, doubled its own Black-White Dissimilarity Index from 0.30 to 0.60 over the same period.

3.4.2 The Teacher Transfer Process in WCPSS

A WCPSS teacher is permitted to seek a teaching position in a different school within the district by completing and submitting the Transfer Request Form (TRF) during the district's annual Employee Transfer Period. This Period often occurred in the spring semester of every school year. During the timeframe of our study, the TRF was converted from a handwritten form to a digital form, either during the 2006, 2007, or 2008 school year. We recovered the TRF for 2005 (Figure 3.1) and for 2009 (Figure 3.2). In terms of functionality and layout, the handwritten and digital versions of the TRF are largely indistinguishable. Listed on the TRF are all schools in the district, organized by the 9 Wake county areas. The school calendar type (e.g., traditional) is indicated on the 2005 TRF but not on the 2009 TRF.

Figure 3.1: Transfer Request Form, 2005

Wake County Public School System
Employee Initiated Transfer Request
2004 - 2005

Employee Initiated Transfer Requests will not be accepted after January 15, 2004.
Any transferee who is offered and accepts a position by May 1, may receive a transfer for the next school year without his/her principal's approval. No transfer will be granted after May 1, without the approval of his/her principal. These deadlines apply to both traditional and year-round schools. ILT 1, 2, 3 teachers are eligible for only one Employee Initiated Transfer while on ILT licensure status. All ILT's are strongly encouraged to refrain from submitting transfer requests until they complete the ILT program. **If you are an ILT and received a transfer in the last two years you are NOT eligible for a transfer this year.**

Last Name First Middle

Home Phone Social Security #

Current School Current Position

Reason (Qualifications):

Requested Position Yrs. of Teaching Experience

Eligible employees are assured of consideration IF a vacancy is anticipated at a traditional and/or year-round school. A copy of this request will be returned to the employee, and placed in his/her personnel file. Employees who are eligible for a transfer will have their name, contact information and certification area(s) sent to the principals of all requested schools. Present Principals will be notified of transfer requests. If a transfer request is approved the employee may send letters of interest, resumes. or emails to principals. Eligible Teachers may attend the Transfer Fair on March 15, 2004 . Additional information will be provided to those teachers approved for a transfer.

CHECK ONLY SCHOOLS YOU WOULD SERIOUSLY CONSIDER.
DO NOT CHECK ALL SCHOOLS.

CENTRAL WAKE	446 ( ) HODGE RD.	451 ( ) HIGHCROFT ELEM	616 ( ) WILBURN (YR)	SOUTHERN WAKE
344 ( ) BROOKS +	451 ( ) FORESTVILLE ELE.	390 ( ) DAVIS DR ELEM (YR)	326 ( ) BAILEYWICK	320 ( ) AVERSBORO
352 ( ) BUGG +	453 ( ) JONES DAIRY (YR)	414 ( ) FARMINGTON +	334 ( ) BRASSFIELD	384 ( ) CREEH RD.
380 ( ) CONN +	454 ( ) HERITAGE (YR)	460 ( ) KINGSWOOD	336 ( ) BRENTWOOD	420 ( ) FUQ- VARINA ELEM
416 ( ) FULLER +	464 ( ) KNIGHTDALE	504 ( ) MORRISVILLE (YR)	396 ( ) DOUGLAS +	447 ( ) HOLLY SPRGS
448 ( ) HUNTER +	480 ( ) LOCKHART	520 ( ) NORTHWOODS	398 ( ) DURANT (YR)	476 ( ) LINCOLN HGHTS +
456 ( ) JOYNER +	544 ( ) ROLESVILLE	522 ( ) OAK GROVE (YR)	415 ( ) FOX RD.	540 ( ) RAND RD.
468 ( ) LACY	584 ( ) WAKE FOREST +	530 ( ) PENNY RD.	440 ( ) GREEN (YR)	560 ( ) SMITH
524 ( ) OLDS +	600 ( ) WENDELL +	523 ( ) OLIVE CHAPEL	442 ( ) HILBURN	576 ( ) VANCE
532 ( ) POE +	632 ( ) ZEBULON +	393 ( ) DILLARD DR.	452 ( ) JEFFREYS GR.	580 ( ) VANDORA SPGS.
536 ( ) POWELL +	593 ( ) WAKEFIELD ELEM	439 ( ) GREENHOPE	469 ( ) LEESVILLE ELEM.	624 ( ) WILLOW SPRGS
548 ( ) ROOT +	WESTERN WAKE	571 ( ) TURNER CREEK (YR)	470 ( ) LEADMINE	570 ( ) TIMBER DR (YR)
572 ( ) UNDERWOOD +	304 ( ) ADAMS (YR)	542 ( ) REEDY CR.	488 ( ) LYNN RD.	449 ( ) HOLLY RIDGE ELEM.
596 ( ) WASHINGTON +	308 ( ) APEX ELEM. +	550 ( ) SALEM ELEM	496 ( ) MILLBROOK ELEM. +	327 ( ) BALLENTINE
620 ( ) WILEY +	328 ( ) BAUCOM	*598 ( ) WEATHERSTONE*	618 ( ) WILDWOOD	494 ( ) MIDDLECREEK
525 ( ) PARTNERSHIP (MOD)	340 ( ) BRIARCLIFF		516 ( ) NORTH RIDGE	626 ( ) YATES MILL
EASTERN WAKE	364 ( ) CARY ELEM.	NORTHERN WAKE	531 ( ) PLEASANT UNION	568 ( ) SWIFT CR.
362 ( ) CARVER (ALT)	376 ( ) COMBS +)	628 ( ) YORK	564 ( ) STOUGH	606 ( ) WESTLAKE ELEM (YR)

CENTRAL WAKE	506 ( ) MOORE SQRE * (MOD)	EASTERN WAKE	551 ( ) SALEM MID	WESTERN WAKE
324 ( ) LONGVIEW (ALT)	NORTHERN WAKE	410 ( ) E. WAKE MID	312 ( ) APEX MID.	391 ( ) DAVIS DR. MID
348 ( ) BROUGHTON HIGH +	360 ( ) CARROLL MID	411 ( ) E. WAKE HIGH	400 ( ) E. CARY MID	394 ( ) DILLARD MID
356 ( ) CARNAGE MID. +	408 ( ) E. MILBROOK +	466 ( ) KNIGHTDALE HIGH	316 ( ) APEX HIGH	*604 ( ) W. CARY*
528 ( ) PHILLIPS HIGH (ALT)	399 ( ) DURANT RD. (YR)	592 ( ) WF/ROLEVILLE MID	318 ( ) ATHENS DR. HIGH	SOUTHERN WAKE
388 ( ) DANIELS MID. +	473 ( ) LEESVILLE HIGH	588 ( ) WF/ROLEVILLE HI	368 ( ) CARY HIGH	404 ( ) E. GARNER
412 ( ) ENLOE HIGH +	500 ( ) MILLBROOK HIGH	636 ( ) ZEBULON MID.	508 ( ) MT. VERNON (ALT)	512 ( ) N. GARNER
472 ( ) LIGON MID. +	552 ( ) SANDERSON HIGH	594 ( ) WAKEFIELD MID	441 ( ) GREENHOPE HIGH	424 ( ) FUQUAY MID
370 ( ) CENTENNIAL + (MOD)	608 ( ) W. MILLBROOK	595 ( ) WAKEFIELD HIGH	607 ( ) W. LAKE MID.(YR)	450 ( ) HOLLY RIDGE MID
562 ( ) SE RALEIGH + (MOD)	471 ( ) LEESVILLE MID	WESTERN WAKE	484 ( ) LUFKIN RD MID. (YR)	428 ( ) FUQUAY HIGH
492 ( ) MARTIN MID. +	444 ( ) HERITAGE MID (YR)	543 ( ) REEDY CR. MID	495 ( ) MIDDLE CR. HIGH	436 ( ) GARNER HIGH

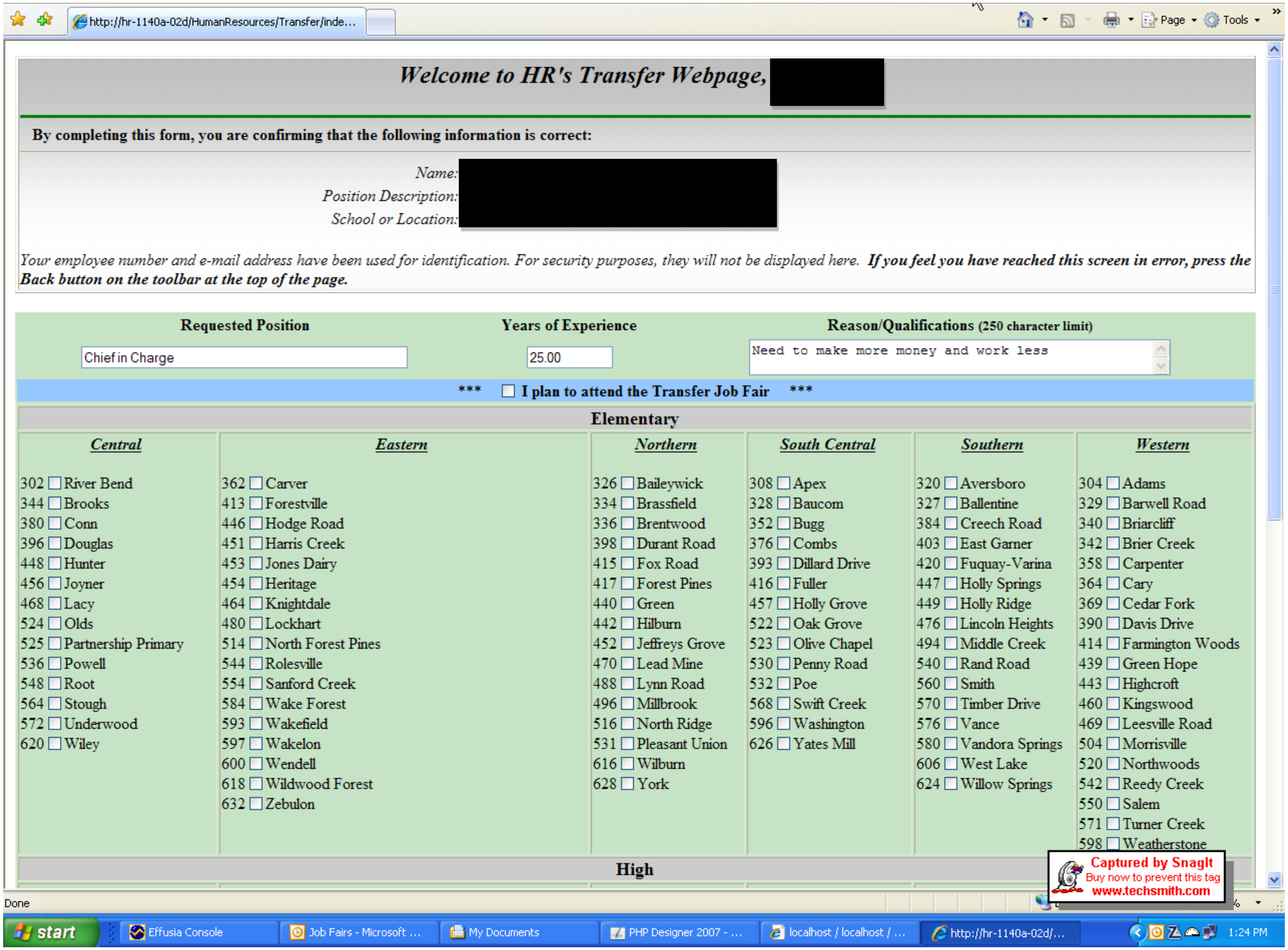
+ MAGNET SCHOOLS (YR) YEAR-ROUND (MOD) MODIFIED CALENDAR (ALT) ALTERNATIVE CALENDAR

ELIGIBILITY DOES NOT GUARANTEE A TRANSFER

(HR USE ONLY) **ELIGIBLE: YES NO REASON (If not eligible):**

HR Admin. Signature **Date:**

Figure 3.2: Sample Transfer Request Form, 2009.



http://hr-1140a-02d/HumanResources/Transfer/inde... Page Tools

Welcome to HR's Transfer Webpage, [redacted]

By completing this form, you are confirming that the following information is correct:

Name: [redacted]
Position Description: [redacted]
School or Location: [redacted]

Your employee number and e-mail address have been used for identification. For security purposes, they will not be displayed here. ***If you feel you have reached this screen in error, press the Back button on the toolbar at the top of the page.***

Requested Position	Years of Experience	Reason/Qualifications (250 character limit)
Chief in Charge	25.00	Need to make more money and work less

*** ☐ I plan to attend the Transfer Job Fair ***

Elementary

Central	*Eastern*	*Northern*	*South Central*	*Southern*	*Western*
302 ☐ River Bend	362 ☐ Carver	326 ☐ Baileywick	308 ☐ Apex	320 ☐ Aversboro	304 ☐ Adams
344 ☐ Brooks	413 ☐ Forestville	334 ☐ Brassfield	328 ☐ Baucom	327 ☐ Ballentine	329 ☐ Barwell Road
380 ☐ Conn	446 ☐ Hodge Road	336 ☐ Brentwood	352 ☐ Bugg	384 ☐ Creech Road	340 ☐ Briarcliff
396 ☐ Douglas	451 ☐ Harris Creek	398 ☐ Durant Road	376 ☐ Combs	403 ☐ East Garner	342 ☐ Brier Creek
448 ☐ Hunter	453 ☐ Jones Dairy	415 ☐ Fox Road	393 ☐ Dillard Drive	420 ☐ Fuquay-Varina	358 ☐ Carpenter
456 ☐ Joyner	454 ☐ Heritage	417 ☐ Forest Pines	416 ☐ Fuller	447 ☐ Holly Springs	364 ☐ Cary
468 ☐ Lacy	464 ☐ Knightdale	440 ☐ Green	457 ☐ Holly Grove	449 ☐ Holly Ridge	369 ☐ Cedar Fork
524 ☐ Olds	480 ☐ Lockhart	442 ☐ Hilburn	522 ☐ Oak Grove	476 ☐ Lincoln Heights	390 ☐ Davis Drive
525 ☐ Partnership Primary	514 ☐ North Forest Pines	452 ☐ Jeffreys Grove	523 ☐ Olive Chapel	494 ☐ Middle Creek	414 ☐ Farmington Woods
536 ☐ Powell	544 ☐ Rolesville	470 ☐ Lead Mine	530 ☐ Penny Road	540 ☐ Rand Road	439 ☐ Green Hope
548 ☐ Root	554 ☐ Sanford Creek	488 ☐ Lynn Road	532 ☐ Poe	560 ☐ Smith	443 ☐ Highcroft
564 ☐ Stough	584 ☐ Wake Forest	496 ☐ Millbrook	568 ☐ Swift Creek	570 ☐ Timber Drive	460 ☐ Kingswood
572 ☐ Underwood	593 ☐ Wakefield	516 ☐ North Ridge	596 ☐ Washington	576 ☐ Vance	469 ☐ Leesville Road
620 ☐ Wiley	597 ☐ Wakelon	531 ☐ Pleasant Union	626 ☐ Yates Mill	580 ☐ Vandora Springs	504 ☐ Morrisville
	600 ☐ Wendell	616 ☐ Wilburn		606 ☐ West Lake	520 ☐ Northwoods
	618 ☐ Wildwood Forest	628 ☐ York		624 ☐ Willow Springs	542 ☐ Reedy Creek
	632 ☐ Zebulon				550 ☐ Salem
					571 ☐ Turner Creek
					598 ☐ Weatherstone

High

Captured by SnagIt
Buy now to prevent this tag
www.techsmith.com

Done

start | Effusia Console | Job Fairs - Microsoft ... | My Documents | PHP Designer 2007 - ... | localhost / localhost / ... | http://hr-1140a-02d/... | 1:24 PM

548 ☐ Root	554 ☐ Sanford Creek	488 ☐ Lynn Road	532 ☐ Poe	560 ☐ Smith	443 ☐ Highcroft
564 ☐ Stough	584 ☐ Wake Forest	496 ☐ Millbrook	568 ☐ Swift Creek	570 ☐ Timber Drive	460 ☐ Kingswood
572 ☐ Underwood	593 ☐ Wakefield	516 ☐ North Ridge	596 ☐ Washington	576 ☐ Vance	469 ☐ Leesville Road
620 ☐ Wiley	597 ☐ Wakelon	531 ☐ Pleasant Union	626 ☐ Yates Mill	580 ☐ Vandora Springs	504 ☐ Morrisville
	600 ☐ Wendell	616 ☐ Wilburn		606 ☐ West Lake	520 ☐ Northwoods
	618 ☐ Wildwood Forest	628 ☐ York		624 ☐ Willow Springs	542 ☐ Reedy Creek
	632 ☐ Zebulon				550 ☐ Salem
					571 ☐ Turner Creek
					598 ☐ Weatherstone

High

Central	***Eastern***	***Northern***	***South Central***	***Southern***	***Western***
324 ☐ Longview	411 ☑ East Wake	500 ☐ Millbrook	316 ☐ Apex	428 ☐ Fuquay-Varina	368 ☐ Cary
348 ☐ Broughton	466 ☐ Knightdale	552 ☐ Sanderson	318 ☐ Athens	436 ☐ Garner	441 ☐ Green Hope
412 ☐ Enloe	588 ☑ Wake Forest-Rolesville		562 ☐ Southeast Raleigh	455 ☐ Holly Springs	473 ☐ Leesville Road
528 ☐ Phillips	595 ☑ Wakefield			495 ☐ Middle Creek	526 ☐ Panther Creek
583 ☐ Wake Early College of	699 ☐ East Wake School of Health Science -				
	700 ☐ East Wake School of Health Science -				
	701 ☑ East Wake School of IT -				
	702 ☐ East Wake School of Arts Ed & Global Studies -				
	703 ☑ East Wake School of Engineering Systems -				

Main Office

Administrative

825 ☐ Project Enlighten

Middle

Central	***Eastern***	***Northern***	***South Central***	***Southern***	***Western***
388 ☐ Daniels	410 ☐ East Wake	360 ☑ Carroll	312 ☑ Apex	404 ☐ East Garner	391 ☐ Davis Drive
472 ☐ Ligon	438 ☐ River Oaks	399 ☐ Durant Road	356 ☐ Carnage	424 ☐ Fuquay-Varina	400 ☐ Reedy Creek
492 ☐ Martin	592 ☐ Wake Forest-Rolesville	408 ☐ East Millbrook	370 ☐ Centennial	450 ☐ Holly Ridge	402 ☐ East Cary
506 ☐ Moore Square	594 ☐ Wakefield	444 ☐ Heritage	394 ☐ Dillard Drive	512 ☐ North Garner	471 ☑ Leesville Road
	601 ☐ Wendell	608 ☑ West Millbrook	484 ☐ Lufkin Road	607 ☐ West Lake	551 ☐ Salem
	636 ☐ Zebulon		508 ☑ Mt Vernon		604 ☐ West Cary

Submit Entries

Clear All Entries

Captured by SnagIt
Buy now to prevent this tag
www.techsmith.com

When viewing the TRF, the teacher observes the name of every school in the district, regardless of whether the school has current vacancies or expects vacancies the following school year. To express interest in transferring to a particular school, the teacher marks the checkbox next to the school's name. The collection of preferred schools might consist of only one school or as many as all schools in the district. There is no limit to the number of preferred schools, although the WCPSS Human Resources (HR) Department encourages teachers to select only schools to which they would seriously consider transferring. We define a teacher as applying to transfer--and thus altering his or her applier status--if the teacher selects at least one school and submits the TRF to the HR Department.

Every year, the collection of TRFs serves as a clearinghouse for the district's appliers. The clearinghouse is helpful to principals, each of whom makes the hiring decisions for the school. During the Employee Transfer Period, the principal is able to identify teachers to contact regarding positions that may be vacated by teachers currently at the principal's school. These vacating teachers are themselves seeking to transfer to other schools. Teachers are also permitted to inquire directly with principals regarding job openings.

We observe in some years of the data that the HR Department held job fairs for teachers seeking to transfer to different schools. The reader can see this indicated on the 2005 and 2008 TRFs. We note that only after a teacher submitted the TRF would the teacher become eligible to attend these job fairs. This fact mitigates the concern that teachers are aware of job openings prior to completing the TRF, and that that awareness influences the lists of preferred schools formed.

3.4.3 Teacher Absence Policy[17]

Two types of absences constitute 81 percent of all of the teacher absences that we observe over the 2004-10 period: vacation days and sick days. Vacation days, or vacation leave, are built into the school calendar. While teachers can accrue vacation leave with more years of experience, they are not permitted to use this form of leave on instructional days. Therefore, vacation leave is of less interest to us than sick leave, whose use in principle is more unpredictable. Teachers accrue without limit sick leave at a rate of 1 sick day per month of work. Therefore, the number of sick days increases in the teacher’s years of experience. If a teacher exhausts this store of sick days in a school year, then the teacher can take sick days at a price of $50 per day, which is for substitute costs. At a teacher’s retirement, unused unpenalized sick days are converted into service credit, which raises the teacher’s state pension benefits.

3.5 Data

We draw upon personnel records for up to 12,480 unique full-time WCPSS classroom teachers observed over the school years 2004-10. Table A5.1 of Appendix 5 reports minimal differences in node-grade reassignment during the years we are able to include in our study (2004-10) and the years under Wake Reassignment Plan implementation (2001-10).

[17] We refer the reader to Clotfelter, Ladd, and Vigdor (2009, pp. 118-121), Gershenson (2016, p. 618), and Ost and Schiman (2017, p. 21) for excellent summaries of the North Carolina teacher absence policy that governs district-level absence policies.

Table 3.1: Descriptive statistics for full-time classroom teachers, Wake County Public School System, 2004-10. Maximum number of teachers:12,480.

	Mean	SD	Count
Key Observable Characteristics			
Proportion Female	0.82	0.38	52955
Proportion Black	0.11	0.31	52955
Proportion Hispanic	0.02	0.13	52955
Proportion Other	0.13	0.34	52955
Proportion White	0.75	0.43	52955
Age on Sep 1	40.22	11.53	52955
Distance from Home to School	25.92	19.12	52955
Years of Experience[1]	12.25	9.38	45882
Proportion with National Board Certification	0.23	0.42	52955
Dependent Variables			
Proportion Resigned for Any Reason	0.05	0.22	52955
Proportion Resigned for Other Opportunities	0.33	0.47	2618
Number of Times Applied to Transfer	0.12	0.32	52955
Number of Sick Days	1.98	5.26	52955

Notes: Other race refers to: Asian, multi-racial, not identified, Native Hawaiian/Pacific Islander.
[1]Available only for the school years 2004 and 2006-10.

Table 3.1 reports summary statistics for WCPSS teachers. Over this time period, most teachers are female, White, about 40-years-old, and live about 26 miles from their schools on average. One-quarter of teachers hold NBPTS certification. For context, we note that WCPSS teachers tend to have more years of experience (12.25) and skew female (82%) and White (73%), relative to NYC teachers in Boyd et al. (2011)'s study (75% female, 62% White, an average of 7.5 years of experience). These differences affirm the idea that our study in the WCPSS context complements the analysis conducted by Boyd et al. (2011) in the NYC context.

3.6 Empirical Strategy

We employ OLS models that control for school and year fixed effects. These models exploit within-school, between-year variation in a school's reassignment status. Because every school serves as its own comparison group, the relevant analytic comparison is within a school and between reassignment and non-reassignment years. The identifying assumption is that a school's reassignment status is plausibly exogenous, conditional on time-invariant school and year unobservables. We bolster this assumption by controlling in our specifications for time-invariant and -varying teacher characteristics.

Our empirical strategy unfolds in four steps. First, we begin by building on Carlson et al. (2020), who found that the Wake Plan meaningfully influenced school-level racial composition inasmuch as Black students were assigned to schools under the Wake Plan that were 38% White, whereas these same schools would have only been 14% White under a standard residence-based assignment plan. We estimate the effect of the Wake Plan on school-level characteristics, which allows us to assess the relative importance of the channels through which reassignment may have operated to affect teacher outcomes. Specifically, we investigate the effects of the Plan on two school-level measures that the Plan explicitly targeted and that we expect it to have influenced: (a) socioeconomic disadvantage, and (b) achievement.

$$(1)\ Y_{st} = \beta_0 + \beta_1 Reassignment_{st} + \sum_{s=1}^{s-1} S_s + \sum_{t=1}^{t-1} T_t + \varepsilon_{ist}.$$

Y denotes for school *s* in year *t* a measure of FRL or achievement, which enters the model as either a log or a level. To construct our measure of FRL, we count the number of FRL students assigned to the school. To construct our measure of achievement, we count the number of students assigned to school *s* in year *t* whose year *t*-1 achievement level was below proficient on one or more end-of-grade or -course tests. *Reassignment* is the independent variable of interest, and we refer to the operationalized construct as a school's *reassignment status*. *Reassignment* denotes an indicator variable equal to 1 for school *s* in year *t* if the school is assigned one or more node-grade units--and thus the students therein--that in year *t*-1 were not assigned to the school. *Reassignment* equals 0 if the list of node-grade units assigned to school *s* in year *t* is identical to the list from school *s* in year *t*-1. Controlling for school fixed effects *S* and year fixed effects *T* allows us to identify the treatment effect using within-school, between-year variation in whether a school is affected by the Wake Reassignment Plan. $\varepsilon$is the stochastic error term. β_1, the coefficient of interest, denotes how much the reassignment of students in certain years to the school affects the outcome, relative to years when students were not reassigned to the school.

Second, we estimate the overall effect of the Wake Reassignment Plan on teacher outcomes. We estimate the following equation:

$$(2)\ Y_{it} = \beta_0 + \beta_1 Reassignment_{st} + \sum_{s=1}^{s-1} S_s + \sum_{t=1}^{t-1} T_t + \beta_2 I_i + \beta_3 D_{it} + \varepsilon_{ist},$$

where *Y* denotes for teacher *i* in year *t* a binary indicator for resignation, a binary indicator for applier status, or sick days coded as a discrete variable. *I* denotes time-invariant observable characteristics for teacher *i*, specifically an indicator for female and separate indicators for Black, Hispanic, and Other (White omitted). *D* denotes "as the crow flies" distance from home to school for teacher *i* in year *t*. All other terms are the same as they were in Equation (1).

Third, we explore the extent to which the treatment effect (i.e., the coefficient estimate on *Reassignment*) varies by two measures of teacher quality. The first measure of teacher quality is years of experience. Relative to more-experienced teachers, those with fewer years of experience are more likely to transfer to other schools (Boyd et al., 2011) and are less likely to take days off from work, because during their short tenure they have not accumulated much time off (Clotfelter, Ladd, and Vigdor 2009). The second measure of teacher quality is National Board for Professional Teaching Standards (NBPTS) certification. National Board Certified Teachers (NBCTs) more than non-NBCTs employ a wider variety of student formative assessments and more effectively use the results to inform student learning (Sato, Wei, and Darling-Hammond 2008). Therefore, we suspect that NBCTs in schools affected by reassignment did not meaningfully differ on the teacher outcomes we study. Across the two reassignment conditions, NBCTs would have taken comparable levels of sick days as they meet the potentially heightened pedagogical demands associated with teaching skill-diverse student populations of the kind assigned to receiving-schools. We also suspect that NBCTs would have been similar in both applier status and resignation likelihood across the two reassignment conditions.

We explore treatment effect heterogeneity in two ways. First, we re-estimate Equation (2) and restrict the analytic sample to NBCTs. Second, to explore how much the treatment effect varies by a teacher's years of experience, we estimate Equation (3):

$$Y_{it} = \beta_0 + \beta_1 Reassignment_{st} + \beta_2 YearsExp_{it} + \beta_3 (Reassignment_{st}\ x\ YearsExp_{it}) + \sum_{s=1}^{s-1} S_s + \sum_{t=1}^{t-1} T_t + \beta_4 I_i + \beta_5 D_{it} + \varepsilon_{ist}$$

where *YearsExp* denotes a discrete variable for years of experience for teacher *i* in year *t*. All other terms from Equation (2) are the same. β_3, the coefficient of interest, denotes the change in the outcome for teacher *i* in year *t* associated with one more year of experience for the teacher when the school is affected by the Wake Reassignment Plan relative to when the school is unaffected by the Plan.

Finally, we explore the possibility that the Reassignment Plan may have had a delayed effect on teacher outcomes. Immediate changes to workplace demands and stressors may not have provided enough information for the teacher to decide whether to leave the job at the end of a reassignment year. Perhaps teachers were more likely to resign or apply to transfer in the school year after reassignment. To explore this possibility, we re-estimate Equations (2) and (3) and lag the independent variable of interest *Reassignment* by one year.

3.7 Results

Table 3.2: The relationship between reassignment under the Wake Reassignment Plan and school characteristics, Wake County Public School System, 2004-10.

	(1)	(2)	(3)	(4)	(5)	(6)
	non-White[1]		FRL[1]		Prior-Year Below Proficient[2]	
	Log	Level	Log	Level	Log	Level
Reassignment$_{st}$	0.04	8.57	0.06**	11.52**	0.01	-3.65
	(0.03)	(7.48)	(0.03)	(5.59)	(0.03)	(4.73)
R^2	0.851	0.901	0.826	0.854	0.952	0.962
Observations	956	956	956	956	580	580

Notes: Outcome is noted in the column header. All specifications control for school FE and year FE. Standard errors in parentheses and clustered on the school. *p<0.1, **p<0.05, ***p<0.01. [1]Estimated over years 2004-10. [2]Estimated over years 2007-10 due to data availability.

Table 3.2 reports the results from the "first stage" (i.e., Equation [1]), in which we regress school-level characteristics on an indicator for the teacher's school's reassignment status and both school and year fixed effects. Reassignment raised a school's non-White share (i.e., the share of Blacks and Hispanics assigned to the school) by a little more than 4 percent (Col. 1), or about 12 non-White students (Col. 2). We note that these coefficient estimates are indistinguishable from zero. As expected, reassignment increased a school's FRL share by a little more than 6 percent (Col. 3), or about 16 students (Col. 4); both estimates are significant at conventional levels. Finally, reassignment did not appear to alter a school's share of students that scored below-proficient on one or more end-of-grade or -course tests during the prior year (Cols. 5 and 6).

Table 3.3: The effect of the Wake Reassignment Plan on teacher outcomes, 2004-10.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Resigned				Applier Status		Sick Days	
	For Any Reason		For Other Opportunities					
	Panel A. All Teachers							
$Reassignment_{st}$	0.00	0.00	-0.04	-0.04	-0.01	-0.01	-0.11**	-0.11**
	(0.00)	(0.00)	(0.03)	(0.03)	(0.01)	(0.01)	(0.05)	(0.05)
R^2	0.011	0.127	0.098	0.132	0.038	0.039	0.060	0.063
Observations	52955	52955	2618	2618	52955	52955	52955	52955
	Panel B. National Board Certified Teachers Only							
$Reassignment_{st}$	-0.00	-0.00	0.03	0.10	-0.00	-0.00	-0.03	-0.02
	(0.00)	(0.00)	(0.12)	(0.12)	(0.01)	(0.01)	(0.09)	(0.09)
R^2	0.019	0.098	0.527	0.571	0.070	0.071	0.101	0.104
Observations	12285	12285	236	236	12285	12285	12285	12285
	Panel C. Heterogeneity by Years of Teaching Experience							
$Reassignment_{st}$ x $YrsExp_{it}$	0.00	0.00	0.00*	0.00**	0.00	0.00	-0.01	-0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
R^2	0.012	0.126	0.216	0.233	0.038	0.039	0.129	0.130
Observations	45882	45882	2189	2189	45882	45882	45882	45882
Covariates	No	Yes	No	Yes	No	Yes	No	Yes

Notes: The outcome is noted in the column header or sub-header. A teacher resigning for other oppunities did so for the following reasons: to stay in the district but in a non-teaching position, to move into a non-teaching position in education, for employment outside of education, to teach in a different NC district, to move to another state or government agency, to continue education or take sabbatical, to teach outside of NC, due to dissatisfaction with teaching, due to dissatisfaction with working conditions, to work in an NC charter school, to work in an NC private school, for a career change. Covariates include female, race (Black, Hispanic, Other, White [omitted]), master's degree or higher, National Board Certification, and distance from home to school. All specifications control for school FE and year FE. Standard errors in parentheses and clustered on the school. Asterisks denote statistical significance: *p<0.1, **p<0.05, ***p<0.01.

Table 3.3 presents our first set of "second stage" results (Equation [2]). Panel A reports coefficient estimates based on an analytic sample that includes all teachers. We document no meaningful effect of reassignment on teacher resignation for any reason (Cols. 1 and 2) or to pursue other professional opportunities (Cols. 3 and 4), in reassignment years compared to non-reassignment years. We also find no meaningful effect on applier status in reassignment years compared to non-reassignment years (Cols. 5 and 6). Finally, we find that school reassignment lowered teacher sick days by about 0.11 days during the reassignment years relative to the non-reassignment years (Cols. 7 and 8), which is significant at conventional levels.

Panel B of Table 3.3 reports coefficient estimates recovered from Equation (2) when estimated over NBCTs only. During reassignment years within the same school, NBCT resignation, applier status, and sick days were no more responsive to reassignment than during non-reassignment years. We acknowledge that our estimates when resignation for other opportunities is the outcome are imprecise because of a very small sample (e.g., Cols. 3 and 4).

Panel C of Table 3.3 reports coefficient estimates recovered from Equation (3), where we interact the teacher's years of experience with the indicator for whether the teacher's school was affected by reassignment in a particular year. Zero on the interaction term coefficient estimate implies that the treatment effect does not vary by teacher years of experience within a school, during reassignment years relative to non-reassignment years. However, we find that within the same school during reassignment years, the likelihood that any of these teachers resigned to pursue other professional opportunities declined by about 9 percentage points, relative to non-reassignment years (Cols. 3 and 4). We also find that the likelihood that a teacher applies to transfer declined by about 1 percentage point during reassignment years relative to non-reassignment years (Cols. 5 and 6). Finally, while the coefficient estimate on $Reassignment_{st}$

implies that sick days increased in years of experience, we also document a small and insignificant offsetting effect as years of experience increases considerably, which is implied by the negative but modest coefficient estimate on the interaction term $Affected_{st}$ x $YrsExp_{it}$ (Cols. 7 and 8).

Table 3.4: The effect of the Wake Reassignment Plan on teacher outcomes, once-lagged effects, 2004-10.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Resigned				Applier Status		Sick Days	
	For Any Reason		For Other Opportunities					
	Panel A. All Teachers							
Reassignment$_{st-1}$	0.00	0.00	0.01	-0.00	0.01	0.01	-0.16**	-0.16**
	(0.00)	(0.00)	(0.03)	(0.03)	(0.01)	(0.01)	(0.06)	(0.06)
R^2	0.006	0.122	0.134	0.176	0.046	0.047	0.069	0.071
Observations	40655	40655	1859	1859	40655	40655	45362	45362
	Panel B. Heterogeneity by Years of Teaching Experience							
Reassignment$_{st-1}$ x YrsExp$_{it}$	-0.00	-0.00	0.00	0.00	0.00	0.00	-0.04***	-0.04***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
R^2	0.009	0.127	0.305	0.326	0.048	0.049	0.129	0.130
Observations	35141	35141	1584	1584	35141	35141	39786	39786
Covariates	No	Yes	No	Yes	No	Yes	No	Yes

Notes: The outcome is noted in the column header or sub-header. A teacher resigning for other oppunities did so for the following reasons: to stay in the district but in a non-teaching position, to move into a non-teaching position in education, for employment outside of education, to teach in a different NC district, to move to another state or government agency, to continue education or take sabbatical, to teach outside of NC, due to dissatisfaction with teaching, due to dissatisfaction dissatisfied with working conditions, to work in an NC charter school, to work in an NC private school, for a career change. Covariates include female, race (Black, Hispanic, Other, White [omitted]), master's degree or higher, National Board Certification, and distance from home to school. All specifications control for school FE and year FE. Standard errors in parentheses and clustered on the school. Asterisks denote statistical significance: *p<0.1, **p<0.05, ***p<0.01.

Finally, to address the possibility that teachers may have had a delayed response to reassignment, Table 3.4 reports the effects of once-lagged school reassignment on teacher outcomes. Coefficient estimates in Panels A and B are qualitatively similar to their counterparts in Table 3.3. We note that the coefficient estimates in Panel A of Table 3.4 for sick days (Cols. 7 and 8) are larger in magnitude than in Panel A of Table 3.3, meaning that in the year following reassignment teachers take even fewer sick days on average relative to the reassignment year. In Panel C of Table 3.4, an offsetting effect similar to that observed in Panel C of Table 3.3 emerges for sick days. Within the same school, sick days increase in reassignment years relative to non-reassignment years (Cols. 7 and 8); however, teachers with more years of experience appear to reduce their sick days more still.

3.8 Discussion

We extend the literature on contemporary school desegregation efforts by broadening their collective scope to include what, to the best of our knowledge, are teacher outcomes that have received little attention in the same context. We found that teacher resignation and applier status within a school were largely unresponsive during reassignment years relative to non-reassignment years. Perhaps surprisingly, reassignment appears to have had a negative effect on sick days during reassignment years relative to non-reassignment years. We acknowledge that there are probably several plausible reasons for this result. The explanation most compelling to us, however, is that teachers take fewer sick days during reassignment years because they increase their instructional efforts in response to teaching a student population that is more skill-diverse than ones they had seen during non-reassignment years. This particular explanation is supported by our once-lagged estimates (Table 3.4, Panel A, Cols. 7 and 8), which show that in the year after reassignment teachers reduce sick day consumption even more than they did during the reassignment year. One explanation here is that in the year after reassignment teachers have adjusted somewhat to delivering instruction to the more skill-diverse student population, therefore requiring fewer sick days. Future research should explore this and other mechanisms.

We acknowledge two important limitations of our study. First, our data restrict us to measuring teacher exposure to changes in *school-level* sociodemographic characteristics. Future studies might attempt to affirm or refine our results using a more granular measure of exposure, perhaps by focusing on grade- or classroom-level, reassignment-induced changes in sociodemographic characteristics. Second, perhaps reassignment-induced changes in school-level sociodemographic characteristics have small to null effects on resignation and applier status because such changes were quite modest. One related goal of future studies might be to catalog the magnitude of these changes across contexts where contemporary school desegregation plans are implemented (e.g., Berkeley, NYC, Wake County). Information of this sort would offer researchers and policymakers a sense of how drastically or modestly these plans alter school-level sociodemographic characteristics.

Our findings reaffirm the benefits of school desegregation. To the extent that the Wake Reassignment Plan is comparable to school desegregation plans in other contexts, our study suggests that these plans have little impact on teacher outcomes of concern to policymakers and school system administrators addressing issues of teacher turnover. In fact, our study suggests that teachers respond positively to the consequences of school desegregation, one of which is the potential skill-diversification of student populations. Replication of the current study in other contexts of school desegregation are warranted.

REFERENCES

Bassok, Daphna, and Sean F. Reardon. 2013. "'Academic Redshirting' in Kindergarten: Prevalence, Patterns, and Implications." *Educational Evaluation and Policy Analysis* 35 (3): 283–97. http://www.jstor.org/stable/43773434.

Buckles, Kasey S., and Daniel M. Hungerman. 2013. "Season of Birth and Later Outcomes: Old Questions, New Answers." *The Review of Economics and Statistics* 95 (3): 711–24. https://doi.org/10.1162/REST_a_00314.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik. 2017. "Rdrobust: Software for Regression-Discontinuity Designs." *The Stata Journal: Promoting Communications on Statistics and Stata* 17 (2): 372–404. https://doi.org/10.1177/1536867X1701700208.

Clemens, Nathan H., Kelsey Ragan, and Christopher Pricket. 2018. "Predictive Validity." In *The Sage Encyclopedia of Educational Research, Measurement, and Evaluation*, edited by Bruce B. Frey. Los Angeles: SAGE Reference.

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31–47. https://doi.org/10.1257/aer.97.2.31.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78 (3): 883–931. https://doi.org/10.3982/ECTA6551.

Datar, Ashlesha. 2006. "Does Delaying Kindergarten Entrance Give Children a Head Start?" *Economics of Education Review* 25 (1): 43–62. https://doi.org/10.1016/j.econedurev.2004.10.004.

Deming, David, and Susan Dynarski. 2008. "The Lengthening of Childhood." *Journal of Economic Perspectives* 22 (3): 71–92.

Dhuey, Elizabeth, David Figlio, Krzysztof Karbownik, and Jeffrey Roth. 2019. "School Starting Age and Cognitive Development." *Journal of Policy Analysis and Management* 38 (3): 538–78. https://doi.org/10.1002/pam.22135.

Dhuey, Elizabeth, and Stephen Lipscomb. 2010. "Disabled or Young? Relative Age and Special Education Diagnoses in Schools." *Economics of Education Review* 29 (5): 857–72. https://doi.org/10.1016/j.econedurev.2010.03.006.

Elder, Todd E., and Darren H. Lubotsky. 2009. "Kindergarten Entrance Age and Children's Achievement: Impacts of State Policies, Family Background, and Peers." *The Journal of Human Resources* 44 (3): 641–83. http://www.jstor.org/stable/20648912.

Fuchs, Lynn S., Douglas Fuchs, Donald L. Compton, Joan D. Bryant, Carol L. Hamlett, and Pamela M. Seethaler. 2007. "Mathematics Screening and Progress Monitoring at First Grade: Implications for Responsiveness to Intervention." *Exceptional Children* 73 (3): 311–30. https://doi.org/10.1177/001440290707300303.

Goodman, Kenneth S., ed. 2006. *The Truth about DIBELS: What It Is, What It Does*. Portsmouth, NH: Heinemann.

Hemelt, Steven W., and Rachel B. Rosen. 2016. "School Entry, Compulsory Schooling, and Human Capital Accumulation: Evidence from Michigan." *The B.E. Journal of Economic Analysis & Policy* 16 (4). https://doi.org/10.1515/bejeap-2015-0219.

Hill, Carolyn J., Howard S. Bloom, Alison Rebeck Black, and Mark W. Lipsey. 2008. "Empirical Benchmarks for Interpreting Effect Sizes in Research." *Child Development Perspectives* 2 (3): 172–77. https://doi.org/10.1111/j.1750-8606.2008.00061.x.

Jenkins, Joseph R., Roxanne F. Hudson, and Evelyn S. Johnson. 2007. "Screening for At-Risk Readers in a Response to Intervention Framework." *School Psychology Review* 36 (4): 582–600.

Lubotsky, Darren, and Robert Kaestner. 2016. "Do `Skills Beget Skills'? Evidence on the Effect of Kindergarten Entrance Age on the Evolution of Cognitive and Non-Cognitive Skill Gaps in Childhood." *Economics of Education Review* 53 (August): 194–206. https://doi.org/10.1016/j.econedurev.2016.04.001.

Manzo, Kathleen Kennedy. 2005. "National Clout of DIBELS Test Draws Scrutiny." *Education Week*, 2005.

Shapiro, Anna. 2020. "Over Diagnosed or Over Looked? The Effect of Age at Time of School Entry On Students Receiving Special Education Services." EdWorkingPapers.com. https://www.edworkingpapers.com/ai20-259.

Strauss, Valerie. 2007. "DIBELS Test: A Question Of Validity," March 26, 2007. http://www.washingtonpost.com/wp-dyn/content/article/2007/03/25/AR2007032501082.html.

Tourangeau, Karen, Christine Nord, Thanh Lê, Alberto G. Sorongon, and Michelle Najarian. 2009. "Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K): Combined User's Manual for the ECLS-K Eighth-Grade and K-8 Full Sample Data Files and Electronic Codebooks. NCES 2009-004." Washington, D.C: National Center for Education Statistics. https://files.eric.ed.gov/fulltext/ED511826.pdf.

APPENDIX 1: CHAPTER 1, APPENDIX A

Table A1.1: Example test schedules, Dewitt Elementary School, Arkansas, 2017-2019.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
NCES School ID	Grade	ACT Aspire Test Subject	School Year	Start Date	End Date	Precise Test Date
050000100218	3	English	2017	4/18/2017	4/19/2017	0
050000100218	3	English	2018	4/30/2018	5/3/2018	0
050000100218	3	English	2019	5/2/2019	5/2/2019	1
050000100218	3	Math	2017	4/18/2017	4/19/2017	0
050000100218	3	Math	2018	4/30/2018	5/3/2018	0
050000100218	3	Math	2019	4/30/2019	4/30/2019	1
050000100218	3	Reading	2017	4/18/2017	4/19/2017	0
050000100218	3	Reading	2018	4/30/2018	5/3/2018	0
050000100218	3	Reading	2019	4/29/2019	4/29/2019	1
050000100218	3	Science	2017	4/18/2017	4/19/2017	0
050000100218	3	Science	2018	4/30/2018	5/3/2018	0
050000100218	3	Science	2019	5/1/2019	5/1/2019	1
050000100218	4	English	2017	4/24/2017	4/25/2017	0
050000100218	4	English	2018	4/23/2018	4/26/2018	0
050000100218	4	English	2019	5/2/2019	5/2/2019	1
050000100218	4	Math	2017	4/24/2017	4/25/2017	0
050000100218	4	Math	2018	4/23/2018	4/26/2018	0
050000100218	4	Math	2019	4/30/2019	4/30/2019	1
050000100218	4	Reading	2017	4/24/2017	4/25/2017	0
050000100218	4	Reading	2018	4/23/2018	4/26/2018	0
050000100218	4	Reading	2019	4/29/2019	4/29/2019	1
050000100218	4	Science	2017	4/24/2017	4/25/2017	0
050000100218	4	Science	2018	4/23/2018	4/26/2018	0
050000100218	4	Science	2019	5/1/2019	5/1/2019	1
050000100218	5	English	2017	4/26/2017	4/27/2017	0
050000100218	5	English	2018	4/16/2018	4/19/2018	0
050000100218	5	English	2019	5/2/2019	5/2/2019	1
050000100218	5	Math	2017	4/26/2017	4/27/2017	0
050000100218	5	Math	2018	4/16/2018	4/19/2018	0
050000100218	5	Math	2019	4/30/2019	4/30/2019	1
050000100218	5	Reading	2017	4/26/2017	4/27/2017	0
050000100218	5	Reading	2018	4/16/2018	4/19/2018	0
050000100218	5	Reading	2019	4/29/2019	4/29/2019	1

050000100218	5	Science	2017	4/26/2017	4/27/2017	0
050000100218	5	Science	2018	4/16/2018	4/19/2018	0
050000100218	5	Science	2019	5/1/2019	5/1/2019	1

Note: Arkansas Department of Education Data Center.

Table A1.2: ACT Aspire Tests relative to active-shooter drill, grades 3-8 in the event-study ACT Sample, Arkansas Public Schools, 2016-2019.

	(1)	(2)	(3)	(4)	(5)	(6)
	ACT Aspire Test Relative to Active-Shooter Drill					
	Switchers		Test Always Before Drill		Test Always After Drill	
	Mean	SD	Mean	SD	Mean	SD
Panel A. School-Grade Characteristics						
Percent Black	17.58	25.53	14.16	25.21	20.99	27.44
Percent Hispanic	11.19	14.32	9.57	10.56	13.46	16.47
Percent White	66.55	27.73	71.95	25.73	60.40	30.94
Percent Proficient or Above on ACT Aspire Tests	49.10	18.45	50.09	20.00	50.07	20.23
Panel B. School Characteristics						
Percent Economically Disadvantaged	67.65	17.76	71.13	14.75	66.33	18.95
Attendance Rate	94.94	1.59	94.66	1.46	94.59	1.91
Student-Teacher Ratio	12.25	3.71	10.55	3.51	12.73	3.72
Percent Grades 3-5	57.82	-	60.63	-	64.31	-
Percent Grades 6-8	42.18	-	39.37	-	35.69	-
Percent Grades 9-10	0.00	-	0.00	-	0.00	-
Percent in Cities	20.85	-	3.15	-	28.43	-
Percent in Suburbs	12.09	-	0.79	-	12.81	-
Percent in Towns	20.10	-	25.20	-	13.48	-
Percent in Rural Areas	46.96	-	70.87	-	45.29	-
N School-Grade-Subject	1861		508		4615	
N School-Grade	476		127		1165	
N School	178		49		435	

Note: The unit of observation is the school-grade-subject-year. Percent proficient or above on ACT Aspire Tests is calculated across years for the same school-grade-subject and weighted by the number of test-takers in the school-grade-subject-year cell. Subject Columns (1) and (2) refer to school-grade-subject observations for which testing always occurs before the drill. Columns (3) and (4) refer to school-grade-subject observations for which testing always occurs after the drill. Columns (5) and (6) refer to school-grade-subject observations for which the test occurs before the drill in at least one year and after the drill in at least one year.
Source: Arkansas Department of Education Data Center

Table A1.3: Differences in mean school-grade and school characteristics, school-grade-subject-year observations, grades 3-8, Arkansas public schools, 2016-2019.

	(1) Full Sample	(2) Active-Shooter Drill Exposure Group	(3) Fire Drill Exposure Group	(4) Diff: (2)-(3)
Panel A. School-Grade Characteristics				
Percent Black	21.91	19.35	22.27	-2.92***
Percent Hispanic	13.50	11.64	13.76	-2.12***
Percent White	59.29	64.22	58.60	5.62***
Percent Proficient or Above on ACT Aspire	50.70	50.04	50.79	-0.75
Panel B. School Characteristics				
Percent Economically Disadvantaged	66.40	67.93	66.18	1.75***
Attendance Rate	94.48	94.28	94.51	-0.23***
Student-Teacher Ratio	12.94	12.62	12.98	-0.36***
Proportion in Cities	0.31	0.17	0.33	-0.17***
Proportion in Suburbs	0.12	0.19	0.11	0.08***
Proportion in Towns	0.14	0.13	0.14	-0.01
Proportion in Rural Areas	0.42	0.51	0.41	0.1***
N School-Grade-Subject-Year	15672	1926	13746	

Note: The unit of observation is the school-grade-subject-year. Panel (A) reports school-grade characteristics; the unit of observation is the school-grade-year. Panel (B) reports school characteristics; the unit of observation is the school-year. Column (1) reports means for the analytic sample used in the "difference-in-differences" model. Column (2) reports means for observations that tested before or after an active-shooter drill. Column (3) reports means for observations that tested before or after a fire drill. Column (4) reports the results of a two-tailed t-test of the difference in means. Percent proficient or above on ACT Aspire tests is calculated across subjects for the same school-grade and weighted by the number of test-takers in the school-grade-subject-year cell. Subject refers to the ACT Aspire test: English, reading, math, or science. Asterisks denote statistical significance: *p<0.05 **p<.01 ***p<.001.
Source: Arkansas Department of Eduation Data Center

APPENDIX 2: CHAPTER 1, APPENDIX B

Table A2.1: Robustness checks for attendance analysis estimates (Table 1.2) excluding school-year-quarter records associated with 0-minute active-shooter drills.

	(1)	(2)	(3)	(4)	(5)	(6)
	Exclude 2016 0-minute Active-Shooter Drill			Exclude Schools Ever with 2016 0-Minute Active-Shooter Drill		
	All Observations	Exclude End-of-Quarter Shooter Drills		All Observations	Exclude End-of-Quarter Shooter Drills	
		Exclude Last 1 Week	Exclude Last 2 Weeks		Exclude Last 1 Week	Exclude Last 2 Weeks
ShooterDrill=1	-0.165**	-0.152*	-0.146*	-0.173*	-0.158*	-0.150
	(0.082)	(0.082)	(0.084)	(0.090)	(0.089)	(0.092)
Non-Shooter-Drill Quarter Mean Attendance Rate	94.47	94.48	94.50	94.50	94.50	94.50
R Squared	0.775	0.777	0.777	0.766	0.767	0.768
N School-Year-Quarter	7000	6801	6659	6452	6275	6141
N School-Year	1756	1756	1756	1619	1619	1619
N School	738	738	738	685	685	685

Note: Refer to Table 1.3 notes. Columns (1)-(3) exclude school-year-quarter records for schools that in 2016 reported 0-minute active-shooter drills. Columns (4)-(6) exclude all school-year-quarter records in every year if the school reported in 2016 0-minute active-shooter drills. Columns (2) and (5) exclude (a) school-year-quarter observations in quarters 2, 3, and 4 if active-shooter drill exposure occurred during the last week at the end of quarters 1, 2, or 3; and (b) school-year-quarter observations in quarters 1, 2, and 3 if active-shooter drill exposure occurred during the last week at the end of quarters 1, 2, and 3. Columns (3) and (6) extend the exclusion criteria to the last 2 weeks. Asterisks denote statistical significance: *p<0.1 **p<.05 ***p<.01.
Source: Arkansas Department of Eduation Data Center

Table A2.2: Robustness checks for active-shooter drill exposure and ACT Aspire Proficiency Rates (Panel A of Table 1.4).

	(1)	(2)	(3)	(4)
	Exclude 2016 0-minute Active-Shooter Drill		Exclude Schools Ever with 2016 0-Minute Active-Shooter Drill	
	Grades 3-5	Grades 6-8	Grades 3-5	Grades 6-8
Panel A. English				
TestWeek=Week after Drill	-0.30	0.51	-0.30	0.51
	(2.09)	(1.96)	(2.09)	(1.96)
Mean Proficiency Rate Test Week before Drill	74.03	74.46	74.03	74.46
R Squared	0.89	0.91	0.89	0.91
N School-Grade-Year	2392	1401	2392	1401
N School-Year	1106	915	1106	915
N School	449	378	449	378
Panel B. Reading				
TestWeek=Week after Drill	-2.68	-0.29	-2.68	-0.29
	(1.83)	(2.28)	(1.83)	(2.28)
Mean Proficiency Rate Test Week before Drill	40.19	41.88	40.19	41.88
R Squared	0.91	0.93	0.91	0.93
N School-Grade-Year	2389	1406	2389	1406
N School-Year	1104	916	1104	916
N School	449	380	449	380
Panel C. Math				
TestWeek=Week after Drill	-0.64	-4.52**	-0.64	-4.52**
	(2.24)	(2.21)	(2.24)	(2.21)
Mean Proficiency Rate Test Week before Drill	52.26	46.79	52.26	46.79
R Squared	0.91	0.93	0.91	0.93
N School-Grade-Year	2394	1402	2394	1402
N School-Year	1105	913	1105	913
N School	449	378	449	378

Panel D. Science

TestWeek=Week after Drill	-0.06	-3.76*	-0.06	-3.76*
	(1.95)	(1.95)	(1.95)	(1.95)
Mean Proficiency Rate Test Week before Drill	42.30	40.76	42.30	40.76
R Squared	0.92	0.94	0.92	0.94
N School-Grade-Year	2388	1402	2388	1402
N School-Year	1101	913	1101	913
N School	449	378	449	378

Note: Refer to Table 1.4 notes for Panel A. Asterisks denote statistical significance: *p<0.1 **p<.05 ***p<.01.
Source: Arkansas Department of Eduation Data Center

Table A2.3: Robustness checks for estimated effect of active-shooter drill exposure on proficiency rates, using fire drill exposure as a comparison condition (Panel B of Table 1.4).

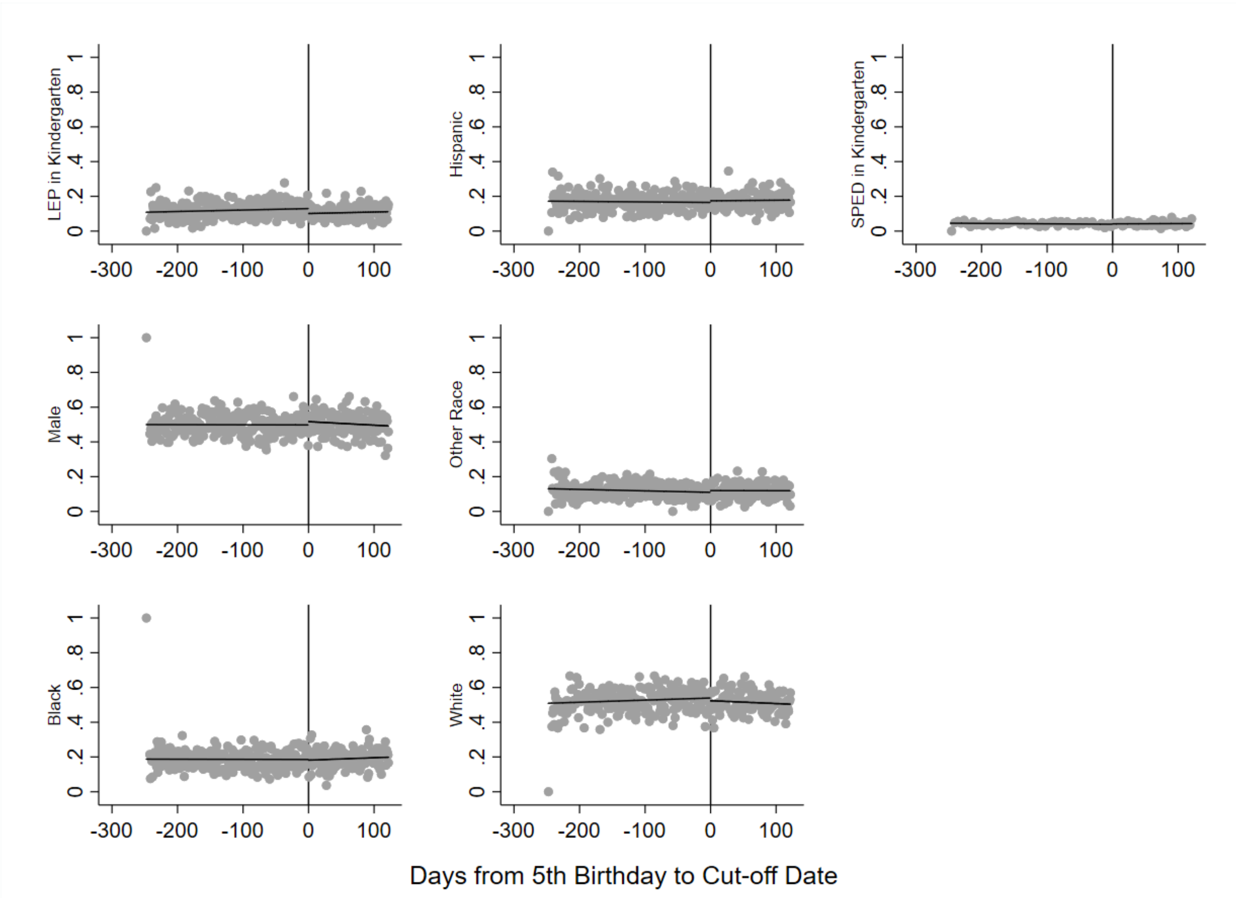
	(1)	(2)	(3)	(4)
	Exclude 2016 0-minute Active-Shooter Drill		Exclude Schools Ever with 2016 0-Minute Active-Shooter Drill	
	Grades 3-5	Grades 6-8	Grades 3-5	Grades 6-8
Panel A. English				
SDrillGroup x DrillExposureIntensity	-0.28	-0.21	-0.40*	-0.24
	(0.22)	(0.15)	(0.24)	(0.16)
Mean Proficiency Rate Test Week before Drill	70.47	74.40	70.45	74.38
R Squared	0.89	0.92	0.89	0.92
N School-Grade-Year	2391	1412	2203	1301
N School-Year	1006	829	934	746
N School	438	385	406	349
Panel B. Reading				
SDrillGroup x DrillExposureIntensity	-0.59***	-0.26	-0.61***	-0.21
	(0.21)	(0.30)	(0.22)	(0.30)
Mean Proficiency Rate Test Week before Drill	38.19	43.87	38.09	43.81
R Squared	0.91	0.93	0.91	0.93
N School-Grade-Year	2383	1404	2195	1293
N School-Year	1003	824	931	741
N School	437	385	405	349
Panel C. Math				
SDrillGroup x DrillExposureIntensity	-0.49**	-0.49**	-0.42*	-0.51**
	(0.22)	(0.24)	(0.24)	(0.25)
Mean Proficiency Rate Test Week before Drill	52.67	48.03	52.81	47.81

R Squared	0.91	0.93	0.91	0.93
N School-Grade-Year	2398	1409	2210	1298
N School-Year	1036	827	1008	744
N School	438	386	496	350
Panel D. Science				
SDrillGroup x DrillExposureIntensity	-0.52**	-0.50*	-0.59***	-0.54*
	(0.22)	(0.26)	(0.21)	(0.28)
Mean Proficiency Rate Test Week before Drill	38.67	43.10	38.59	43.07
R Squared	0.92	0.94	0.92	0.94
N School-Grade-Year	2397	1410	2209	1299
N School-Year	1008	829	936	746
N School	438	386	406	350

Note: Refer to Table 1.4 notes for panel B. Asterisks denote statistical significance: *p<0.1 **p<.05 ***p<.01.
Source: Arkansas Department of Eduation Data Center

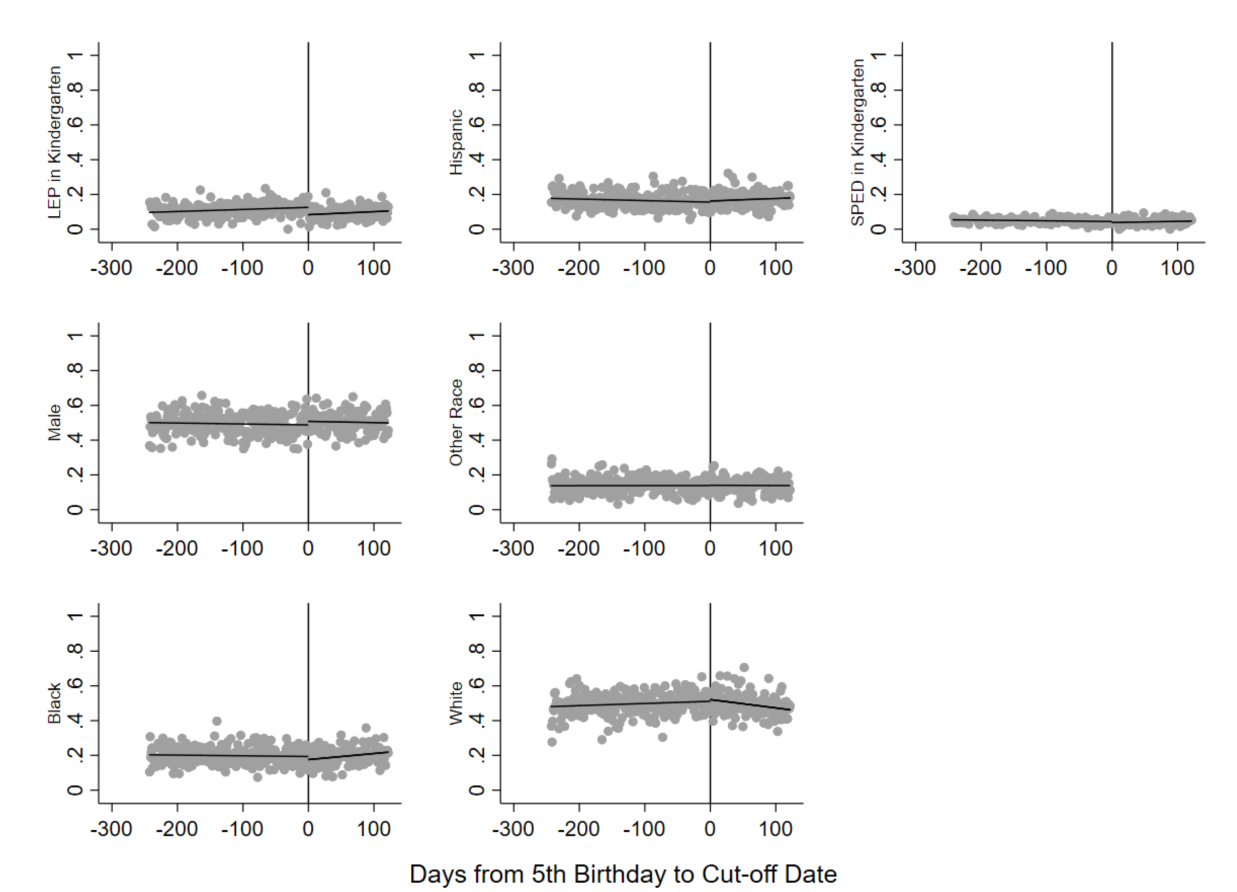
APPENDIX 3: CHAPTER 2, APPENDIX A

Figure A3.1: Covariate smoothness at kindergarten cut-off date, Literacy Sample.



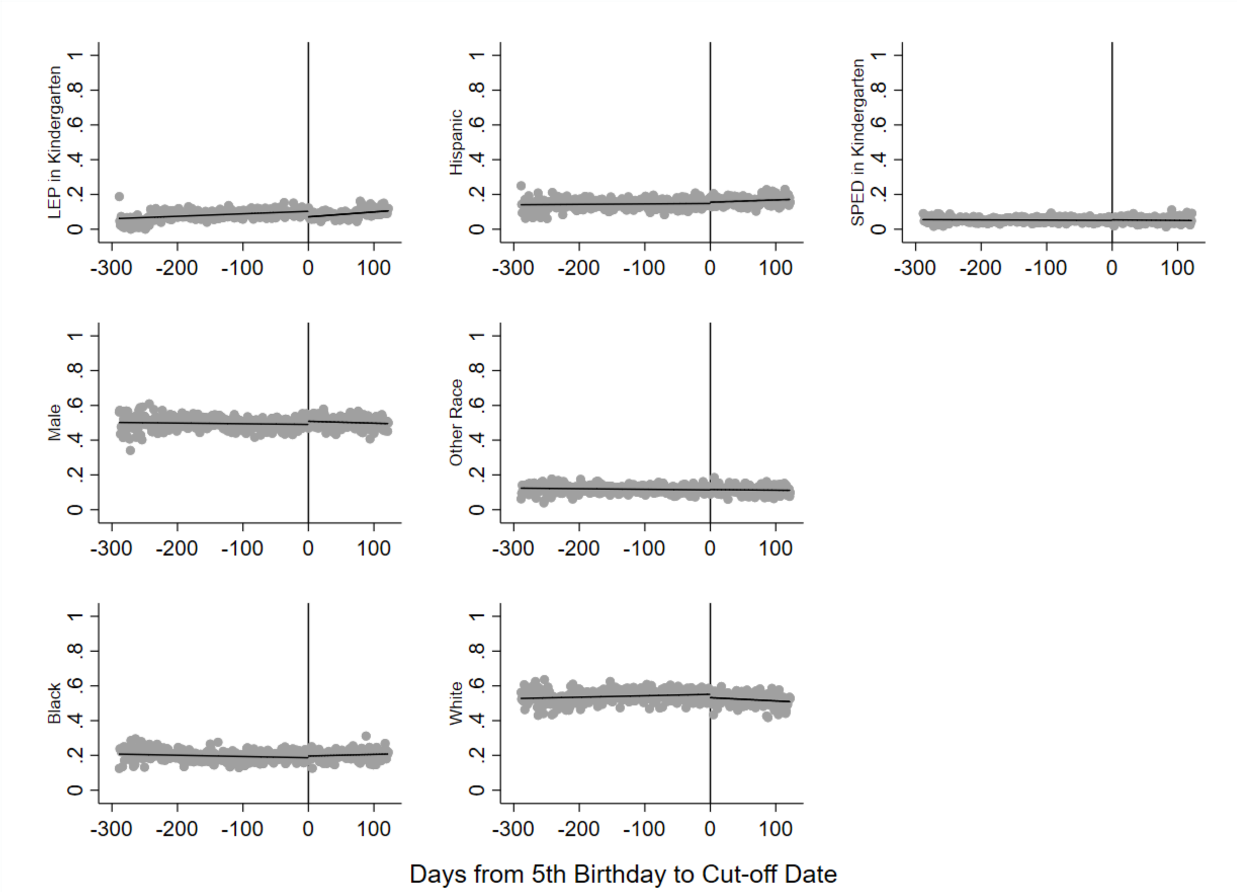
Notes: All covariates are dummy variables. LEP = Limited English Proficient. SPED = Special Education Services. The Literacy Sample consists of kindergarteners entering in the fall of school years 2014-16.

Figure A3.2: Covariate smoothness at kindergarten cut-off date, Behavior Sample.



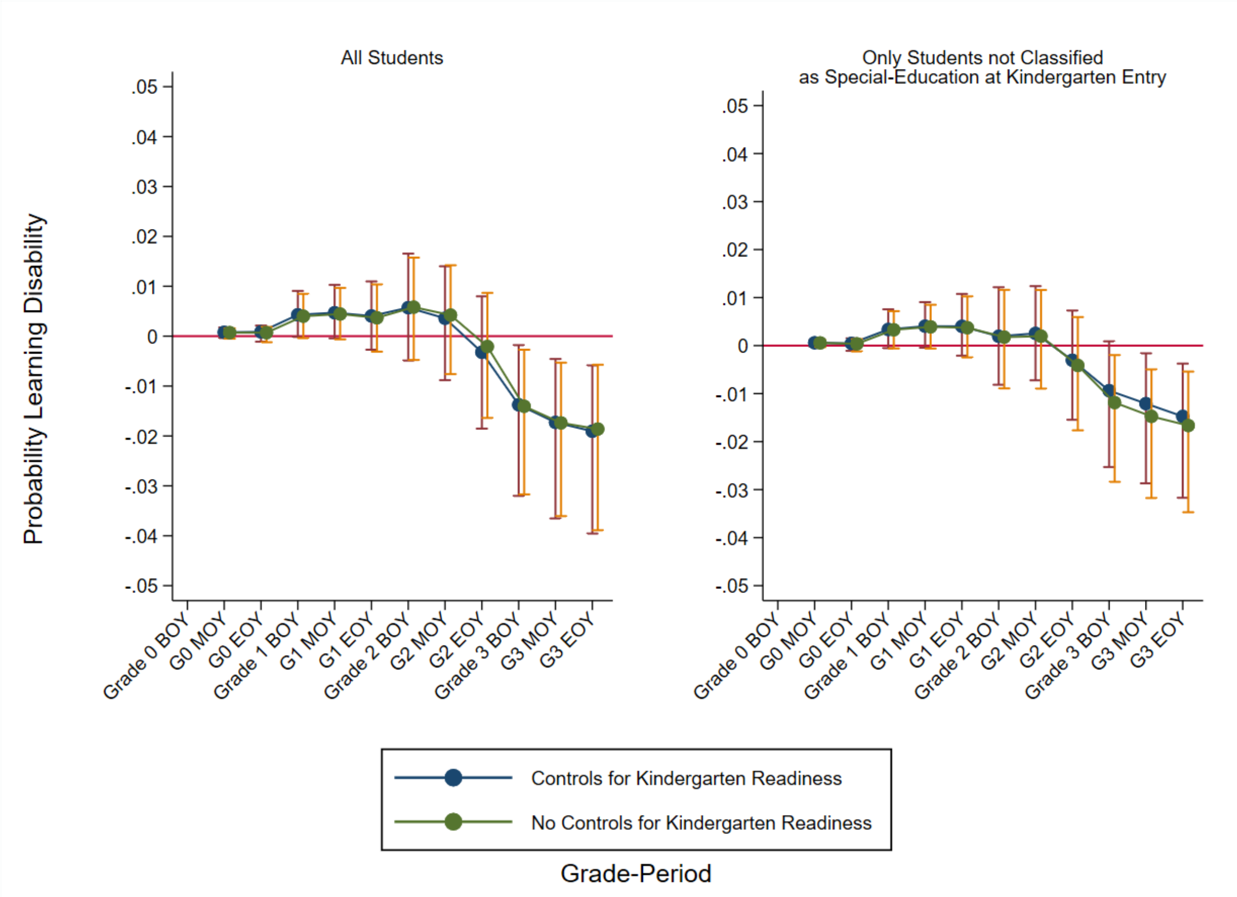
Notes: All covariates are dummy variables. LEP = Limited English Proficient. SPED = Special Education Services. The Behavior Sample consists of kindergarteners entering in the fall of school years 2015-18.

Figure A3.3: Covariate smoothness at kindergarten cut-off date, Disability Sample.



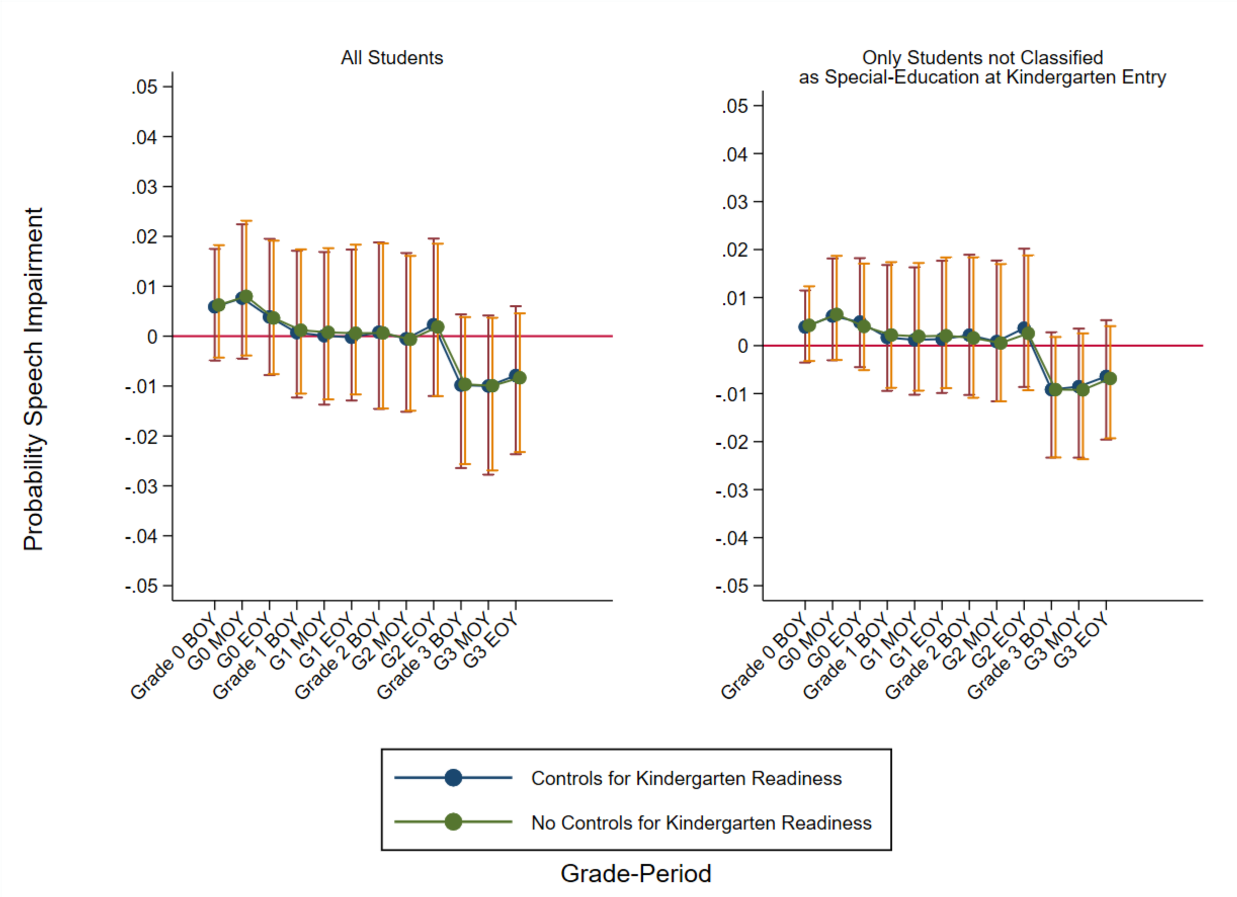
Notes: All covariates are dummy variables. LEP = Limited English Proficient. SPED = Special Education Services. The Disability Sample consists of kindergarteners entering in the fall of school years 2006-16.

Figure A4.4: Probability of learning disability identification, by benchmark period, grades K-3.



Notes: Each point represents the estimated effect of kindergarten enrollment at an older age on the likelihood of being identified with a learning disability, recovered from the non-parametric RD specification estimated over the Disability Sample. Controls for kindergarten readiness include pre-kindergarten attendance and frequency parent reads to child. Additional control variables include: gender; race; limited English proficiency; and school year fixed effects. Grade 0 refers to kindergarten. BOY, MOY, and EOY denote beginning- (months 1-3), middle- (months 4-6), and end-of-year (months 7-9), respectively. Robust 95% confidence intervals. Grade 0 BOY is omitted because a very small number of students are identified with a learning disability at that grade-period, thereby precluding estimation.

Figure A5.5: Probability of speech impairment identification, by benchmark period, grades K-3.



Notes: Each point represents the estimated effect of kindergarten enrollment at an older age on the likelihood of being identified with a speech impairment, recovered from the non-parametric RD specification estimated over the Disability Sample. Controls for kindergarten readiness include pre-kindergarten attendance and frequency parent reads to child. Additional control variables include: gender; race; limited English proficiency; and school year fixed effects. Grade 0 refers to kindergarten. BOY, MOY, and EOY denote beginning- (months 1-3), middle- (months 4-6), and end-of-year (months 7-9), respectively. Robust 95% confidence intervals.

Table A3.1: The effects of older kindergarten enrollment age on DIBELS composite scores, OLS estimates.

	(1)	(2)
Kindergarten		
BOY	0.50	0.42
	(0.07)***	(0.07)***
MOY	0.37	0.29
	(0.06)***	(0.07)***
EOY	0.21	0.14
	(0.07)**	(0.07)
First Grade		
BOY	0.29	0.24
	(0.07)***	(0.07)**
MOY	0.25	0.20
	(0.07)***	(0.07)**
EOY	0.16	0.10
	(0.06)*	(0.06)
Controls for Kindergarten Readiness	No	Yes
Nonparametric Bandwidth Range	64 to 98	76 to 96
Range of Effective Number of Observations	7724 to 9822	7541 to 9015

Notes: The DIBELS composite score is standardized by grade-year-benchmark period to have mean 0 and unit standard deviation. The end-of-grade 3rd grade reading test score is standardized by year to have mean 0 and unit standard deviation. The benchmark periods are: beginning-of-year (BOY), middle-of-year (MOY), and end-of-year (EOY). Controls for kindergarten readiness include: pre-kindergarten attendance; frequency parent reads to child; and measures of kindergarten readiness in oral language, social-emotional skills, and physical coordination. All specifications control for gender, race, limited English proficiency, special-education status, and school year fixed effects. Nonparametric bandwidth range refers to the maximum value of the running variable to the left and to the right of the cut-off date, across all regressions. Range of effective number of observations denotes the maximum number of effective observations used in the estimation, across all regressions. Significance is based on robust p-values: *** p<0.05, ** p<0.01, * p<0.001.

Table A3.2: The effects of older kindergarten enrollment age on the probability of failing to meet or inconsistently meeting Conduct and Work Habits expectations, regression discontinuity estimates.

	(1)	(2)	(3)	(4)
	Conduct		Work Habits	
Kindergarten				
Q1	0.01	0.01	-0.07	-0.07
	(0.03)	(0.03)	(0.03)*	(0.03)*
Q2	-0.01	0.00	-0.07	-0.06
	(0.02)	(0.02)	(0.03)*	(0.03)
Q3	-0.01	0.00	-0.08	-0.07
	(0.02)	(0.02)	(0.03)**	(0.03)*
Q4	-0.03	-0.02	-0.08	-0.06
	(0.03)	(0.03)	(0.02)***	(0.03)*
First Grade				
Q1	0.00	0.00	-0.07	-0.06
	(0.02)	(0.03)	(0.02)**	(0.03)*
Q2	-0.01	-0.01	-0.08	-0.09
	(0.03)	(0.03)	(0.03)**	(0.03)***
Q3	0.00	0.00	-0.07	-0.07
	(0.02)	(0.03)	(0.03)**	(0.03)*
Q4	-0.02	-0.02	-0.06	-0.05
	(0.02)	(0.03)	(0.03)*	(0.03)
Controls for Kindergarten Readiness	No	Yes	No	Yes
Nonparametric Bandwidth Range	73 to 49	70 to 50	78 to 50	72 to 50
Range of Effective Number of Observations	6629 to 8682	5938 to 7789	7496 to 9198	6580 to 7731

Notes: Each coefficient estimate is the conventional local-polynomial non-parametric RD estimate recovered from a distinct regression, calculated using a triangular kernel, heteroskedasticity-robust nearest neighbor variance estimator, and data-driven bandwidth selection procedure described in Calonico et al. (2017). Robust standard errors in parentheses. Conduct is whether the student cooperates with others, respects others, and observes rules and procedures.Work Habits is whether the student uses time wisely, listens carefully, completes assignments, writes legibly, works independently or seeks help when needed, and completes work. Controls for kindergarten readiness include pre-kindergarten attendance and frequency parent reads to child. All specifications control for gender, race, limited English proficiency, special-education status, and school year fixed effects. Nonparametric bandwidth range refers to the maximum value of the running variable to the left and to the right of the cut-off date, across all regressions. Range of effective number of observations denotes the maximum number of effective observations used in the estimation, across all regressions. Significance is based on robust p-values: *** p<0.05, ** p<0.01, * p<0.001.

Table A3.3: The effects of older kindergarten enrollment age on identification with any disability or cognitive disability, regression discontinuity estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Any Disability				Cognitive Disability			
	All Students		Only Students not Classified as Special-Education at Kindergarten Entry		All Students		Only Students not Classified as Special-Education at Kindergarten Entry	
Kindergarten								
BOY	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
MOY	0.01	0.00	0.01	0.01	0.01	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
EOY	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
First Grade								
BOY	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
MOY	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
EOY	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Second Grade								
BOY	-0.01	-0.01	-0.01	0.00	-0.01	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)*	(0.01)*	(0.01)	(0.01)
MOY	-0.01	-0.01	-0.01	-0.01	-0.02	-0.02	-0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)**	(0.01)*	(0.01)	(0.01)
EOY	-0.02	-0.01	-0.01	-0.01	-0.02	-0.01	-0.01	0.00

	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)*	(0.01)	(0.01)	(0.01)
Third Grade								
BOY	-0.02	-0.02	-0.03	-0.02	-0.01	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)*	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
MOY	-0.02	-0.02	-0.02	-0.02	-0.01	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
EOY	-0.01	-0.01	-0.02	-0.01	-0.01	-0.01	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Controls for Kindergarten Readiness	No	Yes	No	Yes	No	Yes	No	Yes
Nonparametric Bandwidth Range	70 to 48	69 to 48	76 to 47	78 to 48	65 to 49	63 to 48	92 to 49	97 to 51
Range of Effective Number of Observations	17955 to 22419	17965 to 21602	18109 to 22038	18557 to 21575	17449 to 22218	16394 to 21073	17372 to 24611	18920 to 26256

Notes: Each coefficient estimate is the conventional local-polynomial non-parametric RD estimate recovered from a distinct regression, calculated using a triangular kernel, heteroskedasticity-robust nearest neighbor variance estimator, and data-driven bandwidth selection procedure described in Calonico et al. (2017). Robust standard errors in parentheses. Cognitive disabilities include: language impairment, mental handicap, intellectual disability, and development delay. The periods are: beginning-of-year (BOY; months 1-3), middle-of-year (MOY; months 4-6), and end-of-year (EOY; months 7-9). Controls for kindergarten readiness include: pre-kindergarten attendance and frequency parent reads to child. All specifications control for gender, race, limited English proficiency, and school year fixed effects. Nonparametric bandwidth range refers to the maximum value of the running variable to the left and to the right of the cut-off date, across all regressions. Range of effective number of observations denotes the maximum number of effective observations used in the estimation, across all regressions. Significance is based on robust p-values: *** p<0.05, ** p<0.01, * p<0.001.

APPENDIX 4: CHAPTER 3, APPENDIX A

Table A4.1: The relationship between third-grade reading test scores and DIBELS composite scores.

	(1)	(2)
Kindergarten		
MOY	0.496***	0.304***
	(0.013)	(0.014)
MOY	0.557***	0.323***
	(0.014)	(0.015)
EOY	0.514***	0.303***
	(0.014)	(0.014)
1st Grade		
BOY	0.527***	0.322***
	(0.014)	(0.014)
MOY	0.650***	0.475***
	(0.012)	(0.013)
EOY	0.716***	0.543***
	(0.012)	(0.014)
Controls	No	Yes
Observations	3753	3753

Notes: The third-grade test score is standardized by grade and year; the DIBELS composite score, grade, year, and benchmark period. Standard errors in parenthses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

APPENDIX 5: CHAPTER 3, APPENDIX B

We observe TRFs for only a subset of the years (2004-10) under the Wake Reassignment Plan (2001-10), which restricts our analysis to only a subset of the Reassignment years. One objection to our study might be that differences in node-grade reassignment exist between the years in our study and the years under the full Wake Plan period. We assess the strength of that objection. Table E.1 reports descriptive statistics on node-grades and schools for the years 2000-10 and 2004-10 separately. The last column on the right denotes the share of all possible records from the years 2000-10 that we can include in our study.

Our analytic sample draws upon nearly 70% of all node-grade-year records, but smaller shares of node-grade records in general, those never reassigned, and those ever reassigned (Panel A). Our analytic sample contains nearly 80% of schools that were ever reassigned during the entire time period under the Wake Plan (Panel B). Importantly, our analytic sample includes many of the schools that ever experienced node-grade reassignment (i.e., through the gain or loss of students due to reassignment) during the Wake Plan period. This finding should allay worries stemming from our use of a subset of the Wake Plan years.

Table A5.1: Node-grade reassignment, counts and schools involved, Wake County Public School System, 2000-10.

	2001-10	2004-10	
	Obs	Obs	Share of 2001-10 Schools (%)
School ever involved in reassignment	143	115	80.4
Elementary	94	85	90.4
Middle	30	29	96.7
High	19	19	100.0

Notes: Combination elementary-middle schools are classified as middle schools; middle-high schools, as high schools. During the 2000-10 period, the Wake County Public School System operated 157 schools: 99 elementary schools, 32 middle schools, and 26 high schools.