Incentives and Productivity in the Economics of Education

by

Uros Petronijevic

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Economics
University of Toronto

Copyright © 2016 by Uros Petronijevic
Abstract

Incentives and Productivity in the Economics of Education

Uros Petronijevic
Doctor of Philosophy
Graduate Department of Economics
University of Toronto
2016

Recent advances in the literature examining the importance of teachers in student learning have led to credible new measures of teacher productivity. Simultaneously, the spread of school-based performance incentives in the United States has fueled a separate literature exploring the way that school performance changes with incentives. These literatures have largely developed independently, as most studies examining teacher effectiveness treat productivity as a fixed trait. This thesis brings the two literatures together, showing how teacher productivity depends on performance incentives.

Chapter 1 examines the effects of school competition on teacher productivity in a setting where public schools face pre-existing accountability incentives. Combining rich student- and teacher-level data in a state that witnessed an expansion of charter schools, I use a difference-in-differences strategy to show that traditional public schools (TPSs) become 0.12 standard deviations more productive when facing charter school competition. Teacher value-added estimates indicate that the school-level improvement is driven entirely by within-teacher performance improvements rather than TPSs hiring more productive teachers.

Chapter 2 explores educator responses to competition in light of prevailing accountability incentives. I propose a model in which teachers differ in their intrinsic motivation, showing the accountability scheme allows unmotivated teachers to exert relatively little effort. Competition then causes school principals to monitor unmotivated teachers more intensively, raising their effort. Consistent with the model, my estimates show that the least-productive teachers drive the average improvement, thus demonstrating that competitive incentives compensate for weaknesses in accountability-program design.
Chapter 3, joint with Hugh Macartney and Robert McMillan, develops an empirical approach to separate the incentive-varying component of teacher productivity – teacher effort – from its incentive-invariant counterpart – teacher ability. The identification strategy uses exogenous variation in incentives from a federal accountability scheme in North Carolina. We find that both effort and ability affect contemporaneous test scores, though with effort having a smaller effect. Both also affect future test scores, with the effect of effort persisting at approximately 25 percent of ability. We show that accountability incentives affect conventional value-added measures, and have implications for the cost effectiveness of policies designed to improve productivity.
Acknowledgements

I am very grateful to my advisors Robert McMillan, Philip Oreopoulos, and Aloysius Siow for their guidance and support throughout this project. Thanks also to Hugh Macartney for helpful discussions during my time as a visiting scholar at Duke University and subsequently.

I would like to thank Natalie Bau, Gustavo Bobonis, Timothy Bond, Elizabeth Dhuey, Trevor Gallen, Michael Gilraine, Yosh Halberstam, Caroline Hoxby, Jonathan Hall, Kirabo Jackson, Kory Kroft, Adam Lavecchia, Joe Mazur, Joshua Murphy, Christopher Ruhm, Ben Sand, Miguel Sarzosa, Andrey Stoyanov, Jim Wyckoff, and seminar participants at the University of Toronto for helpful comments and suggestions.

Support from Canada’s Social Sciences and Humanities Research Council and Ontario’s Graduate Scholarship is gratefully acknowledged. In accordance with the North Carolina Education Research Data Center’s Data Agreement, Hugh Macartney (Duke University) is listed as a coauthor on each chapter of this dissertation. In the standard sense, he and Robert McMillan are coauthors on the third chapter.

Thank you to my parents, Verica and Darko, for your incredible love and support throughout the entire journey and for helping me continue moving forward when I doubted myself. Thank you to my sister, Jelena, for your love, support, and comforting conversations over the years. Thank you to my fiancée, Diane Laranja, for your amazing love and support and for your incredible patience during the stressful times of the economics job market.
Contents

1 School Competition, Accountability, and Teacher Productivity 1
  1.1 Introduction ................................................. 2
  1.2 Data Description and Institutional Details ..................... 6
    1.2.1 Data Description ........................................ 6
    1.2.2 The ABCs of Public Education ............................ 8
    1.2.3 Charter Schools in North Carolina ........................ 11
  1.3 Defining Competition and Estimating its Effects ................ 13
    1.3.1 Measuring School Productivity ............................. 13
    1.3.2 Measuring Teacher Productivity ............................ 15
    1.3.3 Estimating the Effects of Competition on Productivity .... 16
  1.4 The School-Level Effects of Competition ....................... 22
    1.4.1 Robustness Checks ...................................... 23
  1.5 The Teacher-Level Effects of Competition ..................... 26
    1.5.1 Robustness Checks ...................................... 27
  1.6 Conclusion .................................................. 28

2 Competition and Accountability with (Un)Motivated Teachers 42
  2.1 Introduction ............................................... 43
    2.1.1 Relation to Prior Literature .............................. 46
  2.2 A Model of Teachers as Motivated Agents ....................... 49
    2.2.1 Model Setup .............................................. 49
    2.2.2 Effort and Monitoring Decisions without Charter School Competition 52
List of Tables

1.1 Student Characteristics in Schools Serving Grades 3 to 5 from 1997 to 2011 . . . 29
1.2 Schools Serving Grades 3 to 5 from 1997 to 2011 . . . . . . . . . . . . . . . . . . 30
1.3 Teacher Characteristics in Schools Serving Grades 3 to 5 from 1997 to 2011 . . . 31
1.4 Effects of Competition on School Math and Reading Value-Added (VA) . . . . . 32
1.5 Student-Level Regression Results . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
1.6 Effects of Competition on Teacher Math Value-Added (VA) . . . . . . . . . . . . 34

2.1 Heterogeneous Effects of Competition Across TPS Teachers . . . . . . . . . . . . 66
2.2 Heterogeneous Effects with Rank-and-Drop Method . . . . . . . . . . . . . . . . . 67

3.1 Student-Level Summary Statistics . . . . . . . . . . . . . . . . . . . . . . . . . . . 126
3.2 Teacher Performance Variables . . . . . . . . . . . . . . . . . . . . . . . . . . . 127
3.3 The Effects of NCLB Incentives on Teacher Performance . . . . . . . . . . . . . 128
3.4 Tests for Differential Sorting of Students to Teachers in 2003 . . . . . . . . . . . 129
3.5 Parameter Estimates . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 130
List of Figures

1.1 Competitive Effects by Distance from Charter School ......................... 35
1.2 TPS Math and Reading Value-Added Before and After Charter School Opening 36
1.3 Effects of Competition on Enrollment, Staff, and Class-size ..................... 37
1.4 Effects of Competition on Student Body Characteristics .......................... 38
1.4 Effects of Competition on Student Body Characteristics Continued ............. 39
1.5 Teacher Turnover Before and After TPSs Face Competition ....................... 40
1.6 Teacher Math Value-Added Before and After Working in Competition .......... 41
2.1 Heterogenous Effects of Competition on Teacher Value-Added .................. 68
2.1 Heterogenous Effects of Competition on Teacher Value-Added Continued ....... 69
2.2 Heterogenous Effects of Competition on Teacher Salary ........................ 70
2.2 Heterogenous Effects of Competition on Teacher Salary Continued ............. 71
2.3 Heterogenous Effects of Competition on Class Size ............................. 72
2.3 Heterogenous Effects of Competition on Class Size Continued .................. 73
3.1 Inverted-U Response to NCLB .................................................... 131
3.2 Teacher-Year Fixed Effects versus the Proportion of Marginal Students ....... 132
3.3 Incentive-Invariant Ability Distributions ......................................... 132
3.4 Effort Predictions and Effort Distributions in 2003 ............................... 133
3.5 The Relationship Between Teacher Effort and Incentive-Invariant Ability .... 134
3.6 Persistence of Teacher Ability and Baseline Effort in Pre-NCLB Period ........ 134
3.7 Student-Specific Effort Function $e^{nclb}(\hat{y}_{gst} - y_{g,nclb}^T)$ .............. 135
Chapter 1

School Competition, Accountability, and Teacher Productivity
1.1 Introduction

Conventional wisdom among economists deems that competition causes organizations to become more productive. This idea is somewhat controversial in public education, however: not only may the absence of profit maximization impede competitive forces, but it is unclear whether school competition can raise productivity in the current education landscape, where public schools often face strong performance incentives from a variety of accountability programs. These departures from the textbook model call into question how public school productivity will respond to increased competition in such an environment. Do accountability programs lead schools to eliminate inefficiencies, making increased competition superfluous, or do they leave room for competitive forces to cause a further improvement in school productivity?

Addressing this question, this chapter analyzes the effects of school competition in a setting where all schools are held accountable under a sophisticated accountability program prior to facing an increase in competitive pressure. Specifically, I analyze how traditional public schools (TPSs) respond to charter school entry in North Carolina – a state where all TPSs are held accountable under the ABCs of public education, a sophisticated growth-oriented accountability program. Using rich school-administrative data, I am able to track the universe of public school students and teachers over time and accurately measure the productivity of both TPSs and individual teachers. Together with a staggered roll-out of charter schools in North Carolina, these data allow me to estimate how the productivity of TPSs and teachers change when a charter school locates nearby.

To establish a benchmark, I first follow the prior literature and estimate an aggregate effect of increased competition on student tests scores. Applying a difference-in-differences strategy with TPS fixed effects, I show that TPSs are 0.12 standard deviations more productive at raising student math scores when facing competition for student enrollment from a nearby charter school. The rich data allow me to address the two most serious threats to identification of competitive effects: endogenous charter school location decisions and student resorting after competition.

The inclusion of school fixed effects implies that the results are only biased when charter
schools locate in areas based on changes in TPS characteristics that are correlated with changes in TPS productivity. For example, as neighborhoods gentrify slowly, parental pressure could cause gradual improvements in TPS productivity and charter schools may also be more likely to enter such areas because of growth in demand for school quality. In this case, not accounting for the trend in TPS productivity prior to charter school entry would result in upward biased estimates of competitive effects.\footnote{Likewise, if charter schools enter areas where TPS productivity is getting worse, failing to account for the trend would result in downward biased estimates.} I rule out such sources of bias in my analysis by using event-studies to examine past productivity trends directly and by assessing the sensitivity of the results to using a restricted sample of only TPSs that ever faced charter school competition.

Even when positive competitive effects are not driven by the continuation of an improving trend, one may still worry that TPSs have a different group of students before and after an entry event. If unobserved student characteristics are correlated with TPS productivity, and students re-sort across schools after a charter school opens, the estimated productivity effect may capture a student composition effect. I rule out student sorting as a source of bias by estimating nearly identical effects of competition using student-level regressions with student-school fixed effects that use changes in outcomes for the same students attending the same schools to identify the effects of competition. In addition, I estimate no competitive impact on TPS reading productivity, which is consistent with minimal changes in student composition.

The school-level effect I estimate corresponds to 0.025 student-level math score standard deviations, which is within the range estimated by prior studies using longitudinal student-level data and similar identification strategies. When taken in isolation, however, the aggregate result tells us little about the mechanisms by which TPSs achieve these improvements. As teachers are widely recognized to be one of the most important inputs in education,\footnote{This emerging academic consensus (Hanushek, 2011; Bacher-Hicks et al., 2014; Chetty et al., 2014a; Chetty et al., 2014b) is increasingly being reported in the popular press (see, for example, the article titled “How to make a good teacher” in the June 11, 2016 edition of The Economist newspaper).} I focus on teacher quality as a channel for improvement. Indeed, as Hanushek and Rivkin (2003) note, “it would be surprising for competition to exert a substantial effect on students without influencing the quality of teaching.” The question is whether TPSs hire better teachers when facing greater competition or whether existing teachers become more productive. Prior work argues that
competition causes schools to be more selective when making hiring decisions, placing more emphasis on teachers who possess characteristics valued by students (Hoxby, 2002; Hanushek and Rivkin, 2003). Due to data constraints, however, the prior literature has been unable to calculate productivity measures at the teacher level and therefore to assess the extent to which within-teacher productivity improvements drive school-level performance gains.

Using the rich longitudinal data, I find teacher turnover declines at TPSs when a charter school opens nearby, with the fraction of newly-hired teachers falling by 24 percent and there being no change in fraction of teachers leaving. Correspondingly, the school-level effect of competition is almost entirely driven by within-teacher changes in productivity: the average teacher is 0.13 standard deviations more productive after working in a TPS facing competitive pressure from a charter school. I use event-studies at the teacher-level to show that the effect is not driven by teachers sorting into competitive environments based on projected changes in productivity. The estimated effect is also robust to changes in class size and to accounting for improvements in teacher-school match quality that occur after teachers change schools (Jackson, 2013).

These results establish that school competition is an important determinant of measured teacher productivity and shed new light on the effects of competition in education. The vast literature studying school competition typically finds modest aggregate effects of increased competition on student outcomes (Hoxby, 2000; Hoxby, 2003; Sass, 2006; Booker et al., 2008; Bayer and McMillan, 2010; Card et al., 2010; Neilson, 2013; Figlio and Hart, 2014; and Jinnai, 2014). Most prior work does not investigate, however, whether teacher selection or improvement drives aggregate school-level productivity gains. As mentioned, some researchers argue that productivity improvements are more likely to operate through the hiring of better teachers (Hoxby, 2002; Hanushek and Rivkin, 2003), but are unable to test the importance of within-teacher improvement. This chapter provides an empirical test of the within-teacher improvement channel, showing that it explains all of the aggregate school-level competitive effect in North Carolina.

---

3 Jackson (2012) finds a similar result in his analysis of the effects of charter school competition on TPS teacher labor markets.

4 A few studies find no effects (Bettinger, 2005; Bifulco and Ladd, 2006; Rothstein, 2007; Mehta, 2014) or even negative effects (Imberman, 2011).
By establishing the within-teacher improvement effect, this chapter also contributes to the growing literature on the determinants of teacher productivity. Although there still exists some debate (Rothstein, 2014), recent advances in teacher value-added research have shown that there is important heterogeneity across teachers in ability to improve student test scores and that better value-added teachers have lasting effects on their students’ long-run outcomes (Hanushek, 2011; Bacher-Hicks et al., 2014; Chetty et al., 2014a; Chetty et al., 2014b). Researchers and policymakers alike are calling for value-added measures of teacher quality to be used in high-stakes decisions about retention and dismissal (Kane and Staiger, 2014). Surprisingly little is known, however, about the determinants of teacher value-added measures. Indeed, teachers’ unions in Florida, Tennessee, and Texas have argued that a lack of transparency in value-added measures makes them arbitrary and even unconstitutional.

Emerging from this debate is a growing literature concerned with better understanding the determinants of teacher productivity, which shows that peer learning (Jackson and Bruegmann, 2009), task-specific experience (Ost, 2014), teacher-school match quality (Jackson, 2013), school-based accountability incentives (see Chapter 3), colleague mentorship (Papay et al., 2016) teacher-level loss-aversion incentives (Fryer et al., 2012), and the threat of dismissal (Dee and Wyckoff, 2015) all help determine how productive teachers are at raising student test scores. This chapter contributes to that literature by showing that competitive pressure can raise teacher productivity in a clear way, the average teacher moving up approximately five percentile points in the teacher value-added distribution as a result of working in a competitive TPS.

It is worth emphasizing that these effects arise in a state where all schools already face strong accountability incentives from North Carolina’s school-based accountability program – the ABCs of Public Education. The ABCs is a growth-based system that pays financial bonuses to both teachers and principals when schools reach their growth targets. By creating incentives to exert teaching effort throughout the student distribution, growth-based systems provide more uniform incentives for educators than proficiency-count systems like the federal No Child Left Behind (NCLB) Act of 2001, which often result in more attention being given to students who are likely to score near test score proficiency targets (Reback, 2007; Neal and Schanzenbach,
2010; Ladd and Lauen, 2010).

Yet, the documented productivity response to competition in this chapter makes clear that even a program like the ABCs does not provide the incentives necessary to push schools to maximize productivity. In Chapter 2, I use a conceptual model and supporting empirical analyses to show that ABCs incentives allow schools to tolerate within-school disparities in teaching performance, but that competitive incentives cause schools to close the performance gap by requiring more effort from their previously under-performing teachers.

The remainder of the chapter is organized as follows: the next section describes the data and the institutional details in North Carolina. Section 1.3 explains how I estimate the effects of competition on TPSs and teachers. In Section 1.4, I document the effects of competition on school-level productivity, and I show that these effects are driven by within-teacher performance improvements in Section 1.5. Section 1.6 provides concluding remarks.

1.2 Data Description and Institutional Details

1.2.1 Data Description

The data used in this chapter cover the universe of public school students and teachers in North Carolina and are provided by the North Carolina Education Research Data Center (NCERDC). The data contain rich student, teacher, and school characteristics, along with encrypted student and teacher identifiers, which allow me to track and match students and teachers over time and to calculate school and teacher value-added measures. Schools’ longitude and latitude coordinates are also available, allowing me to measure the distance between each TPS and charter school in the state.

Accurately calculating teacher value-added measures requires that I focus on teachers working in third to fifth grade who teach either math or reading, as it is only these teachers who can be accurately matched to their students in the NCERDC data.\(^5\) To be able to relate school-level productivity changes to the productivity changes of individual teachers, I therefore restrict the

\(^5\)To construct a sample of students and teachers who are valid for teacher value-added analyses, I further restrict the sample according to the restrictions in Rothstein (2010). In Rothstein (2010), the analysis sample runs until 2005, whereas my sample runs until 2011, as in Rothstein (2014).
sample to students and teachers who are observed in third, fourth, and fifth grade between the academic years 1997-98 and 2010-11. Table 1.1 presents summary statistics at the student-year level. Column (1) shows summary statistics for the full sample of TPS students. In column (2), I restrict attention to student-year observations that are used in the estimation of teacher value-added. Column (3) considers students attending TPSs located within 5 miles of a charter school serving a grade (from three to five) that overlaps with their TPS, while column (4) shows the characteristics of students attending TPSs that are more than 5 miles away from charter schools. Column (5) restricts the sample to charter school student-year observations.

At the end of each academic year, students write standardized End-of-Grade (EOG) tests in math and reading. Test scores are reported on a developmental scale, but I follow convention in the literature by standardizing scores within grade-year and using standardized scores as my main dependent variable throughout the analysis. Students who are used in the construction of teacher value-added and students who attend TPSs located near charter schools perform better than the average TPS student on these tests. Charter school students score 0.13 standard deviations lower on math tests and 0.03 standard deviations higher on reading tests. Relative to TPSs, charter schools primarily attract under-performing math students but also have lower concentrations of free or reduced-price lunch eligible students and higher concentrations of students with college-educated parents. The average charter school student having a relatively high socioeconomic status (SES) but under-performing in math is consistent with the high-SES parents of disenchanted TPS students being better able to gather the necessary information with which to exercise school choice (Ladd et al. 2015). By providing another option for these students and parents, charter schools may put pressure on TPSs to improve their outcomes – an idea that I investigate in much more detail below.

Table 1.2 presents school-year-level summary statistics across all schools that serve grade three, four, or five from 1997 to 2011. The average charter school has less than half the student enrollment of the average TPS. To get a sense of the grade spans covered by each type of school, I code kindergarten as 0 and calculate the average lowest and highest grade served. The average

---

6Henceforth, I refer to each academic year by its second calendar year. For example, the academic year 2000-2001 is 2001.
Chapter 1. School Competition, Accountability, and Teacher Productivity

TPS serving grades three, four, or five has a grade span of kindergarten to grade five (K-5), while the average charter school has a K-7 span. The table also provides summary statistics for TPSs located at various distances from charter schools. Among the TPSs located within 1 mile of a charter school, nearly 90 percent are in an urban area, compared to 45 percent of TPSs located more than 10 miles away. TPSs within 1 mile also have 0.15 standard deviations lower than average math and reading scores, whereas TPSs within 5 and 10 miles have scores that are much more reflective of the average TPS. Interestingly, TPSs near charter schools have higher math and reading value-added (the construction of these measures is discussed below). For example, the math value-added of TPSs within 5 miles of a charter school is 0.04 student-level standard deviations higher than the average TPS. My empirical strategy below aims to uncover whether this is a causal effect of charter school competition.

Table 1.3 describes the sample of teachers working in schools serving third to fifth grade from 1997 to 2011. TPS teachers are, on average, 4 years more experienced than charter school teachers. They are also 10 percentage points more likely to be female and 8 percentage points less likely to be a racial minority. Reassuringly, teachers who are in the value-added sample are not different than the full sample of TPS teachers in terms of experience, gender, and race. Teachers in TPSs within 5 miles of a charter school have a 0.03 student-level standard deviation higher math value-added and a 0.02 standard deviation higher reading value-added than the average TPS teacher. In Section 1.3, I present a strategy for identifying the causal effect of working in a TPS facing competition on teacher productivity.

1.2.2 The ABCs of Public Education

All TPSs and charter schools in North Carolina are held accountable under the ABCs of public education. North Carolina implemented the ABCs (standing for strong Accountability, teaching the Basics, and maximum local Control) in 1997 and the scheme continued to operate until 2012.\(^7\) Throughout its operation, the scheme was recognized as one of the more sophisticated accountability programs in the United States.\(^8\)

\(^7\)For a detailed description, see http://www.ncpublicschools.org/docs/accountability/reporting/abc/2011-12/backgroundpacket.pdf

\(^8\)See Vigdor (2008) for a detailed description of the ABCs.
The ABCs held schools accountable for student growth, requiring that they sufficiently improve their students’ test scores relative to targets that were functions of students’ prior achievement. In each year, each school faced a set of grade- and subject-specific average growth targets. For both math and reading, the average growth target in each grade depended on the prior math and reading scores of the students currently attending the school. In each grade-subject pair, the target used both math and reading average scores from the prior year to estimate an index of students’ ‘true proficiency’ and the difference between the subject-specific prior average score and the state mean to estimate an index for ‘regression to the mean.’ Targets were designed such that schools with more proficient students were expected to attain higher growth. The targets were also adjusted up or down depending on whether schools were likely to experience positive or negative mean reversion in performance.

At the end of each academic year, when students had written EOG tests in math and reading, the target average growth for each grade-subject pair was subtracted from the realized average growth achieved by all students. The differences between realized and target average growth were then standardized and aggregated across all grade-subject pairs within a school to arrive at a school-level ABCs average growth score. When the school-level score was greater than zero, all teachers and the school principal were given a monetary bonus between $750 and $1500, depending on whether the score was just above zero or greater than the threshold for ‘high’ growth.

The ABCs’ focus on growth sets it apart from proficiency-count programs such as NCLB. By requiring that schools attain targets depending on the prior achievement of their students, the scheme is designed to give all schools a fair chance of passing. In contrast, proficiency-count schemes are arguably more difficult for some schools to satisfy, as these schemes require a given percentage of students within each school to score above an exogenously determined test score target, regardless of the prior achievement of those students. Table 1.2 shows evidence consistent with the superior design of the ABCs, as 78 percent of all TPS-year observations pass the ABCs compared to only 60 percent of observations passing NCLB, which was implemented in North Carolina (and throughout the United States) in 2003.⁹

⁹Some studies have questioned whether the ABCs indeed gives all schools a chance to pass, however, noting
The ABCs’ use of **average** growth targets allows one to derive the school-level targets from the student level, effectively creating a growth target for each student. Seen this way, since each student is required only to sufficiently improve relative to his or her prior performance, the growth component of the ABCs provides educators with incentives to focus on all students throughout the achievement distribution (Macartney *et al.*, 2015). Intuitively, for a given cost of teacher effort, the marginal benefit of teacher effort is the same across all students because an additional unit of growth from any student has the same marginal effect on the probability that the school passes the ABCs. Many studies have shown that this is not the case for proficiency-count programs, which create strong incentives to focus on students who are likely to score around the proficiency threshold (Reback, 2007; Neal and Schanzenbach, 2010; Ladd and Lauen, 2010).

While the ABCs’ focus on student growth makes it relatively sophisticated, the scheme is not without its drawbacks. For example, the method used to adjust targets for mean reversion does not take into account school size, even though mean reversion is likely to be less severe at larger schools. The dependence of the targets on prior scores also makes the targets vulnerable to dynamic distortions through ratcheting behavior on the part of educators (Macartney, 2016). The largest limitation of the scheme, however, is that it creates free-rider incentives across teachers by making individual payoffs contingent on school-level outcomes (Vigdor, 2008; Ahn 2013). Teachers have an incentive to exert less effort than would be the case if they were held individually accountable, as the group-based incentives make teachers incur the full cost of a unit of effort while only increasing the schools’ ABCs score by a fraction proportional to the number of students for which they are responsible.

In theory, then, it is possible for charter school competition to sharpen performance incentives further for TPSs and teachers, compensating in the areas where ABCs incentives are weak. In this chapter, I focus on the aggregate effects of competition, showing that TPS productivity indeed improves when a TPS enters competition with a charter school and that the school-level productivity effect is driven by a given group of teachers improving performance rather than that schools with more proficient students tend to be more likely to pass the ABCs (Ladd and Walsh, 2002; Vigdor, 2008).
TPSs hiring new, more productive teachers. In Chapter 2, I aim to better understand the interaction between ABCs incentives and competitive forces. There, I set out a principal-agent model showing that ABCs group-based incentives result in within-school disparities in teacher effort, but that competitive forces compensate by creating incentives for schools to demand more effort from initially unproductive teachers, thereby closing the gap in performance across teachers and raising school-level productivity. The following subsection describes the threat charter schools pose to TPSs in more detail.

1.2.3 Charter Schools in North Carolina

North Carolina passed legislation allowing charter schools in 1996 (North Carolina General Statute 115C-238.29) and the first charter schools began operating in 1998. The legislation originally allowed no more than 100 schools state-wide, and no more than five new charter openings in any district in a single year. Charter schools rolled out across the state in a staggered fashion, with approximately 30 schools opening in each year from 1998 to 2000. By 2001, North Carolina had almost reached its cap, with 90 charter schools operating across the state.\textsuperscript{10}

Charter schools are encouraged to be innovators in education and are given more freedom with respect to curriculum content and teaching methods than TPSs. They are funded, however, with local, state, and federal tax dollars and are subject to the same accountability programs as all public schools in the state.\textsuperscript{11} Any private non-profit organization can file an application to open a charter school. In North Carolina, an organization must establish a chartering contract with either the local education agency (LEA) in which the school intends to locate, one of the sixteen campuses of the University of North Carolina system, or the State Board of Education, with the State Board being required to give final approval of all contracts. Each application is typically initiated 18 months prior to the proposed opening date, and if the application is successful, the board of the non-profit organization becomes the legal entity responsible for the

\textsuperscript{10}The legislation was updated in 2011 to remove the 100-school cap and to relax some of the laws governing charter operation. North Carolina currently has 158 charter schools operating state-wide.

\textsuperscript{11}A summary of how charters are funded and the flexibility they have with these funds is available here: http://www.ncpublicschools.org/docs/fbs/resources/data/highlights/2004highlights.pdf
As part of the application, charter schools are required to make projections about student enrollment for the next two to five years, explain how they intend to market themselves to potential students and the general public, and estimate the districts from which they will draw most of their students. Since charters do not have catchment areas, their enrollments are entirely driven by parents exercising school choice. These institutional details suggest that charter school location decisions are unlikely to be random. In particular, organizations considering opening a charter school are likely to carefully consider the school’s location, their target student demographic, and the number of students they anticipate enrolling.

Charter school market entry may also generate a response from TPSs, which may take certain measures in an effort to keep their students from leaving to the new competitor. TPSs and charter schools both receive state allotments based on the dollars per average daily membership (ADM) in the LEA in which they are located. Funding for new charters schools is based on the dollars per ADM in the LEA in which each of their students would be enrolled had they not attended the charter school. Importantly, state funding is delivered on a per-student basis and travels with students when they transfer from a TPS to a charter school. TPSs then have an incentive to prevent students from transferring to charter schools in order to avoid a drop in aggregate funding. When TPS students value educational quality and charter schools can offer a level of quality comparable to what TPS students currently receive, charter school competition may cause TPSs to eliminate inefficiencies and become more productive at raising student test scores.

When considering location, their target student demographic, and the number of students they anticipate enrolling, charter school founders are also likely to forecast how local TPSs will respond once they enter the market and start competing for students. Given the careful consideration that may go into choosing charter school locations, estimating the effects of charter school competition on TPS productivity requires an identification strategy that accounts for endogenous charter school location decisions. After describing how I measure TPS and

---

individual teacher productivity, I develop such a strategy in the following section.

1.3 Defining Competition and Estimating its Effects

In this section, I first explain how I construct value-added productivity measures for both TPSs and teachers. With these measures in hand, I then explain how I define competition between TPSs and charter schools and how I estimate the effects of charter school competition on the value-added of TPSs and teachers.

1.3.1 Measuring School Productivity

I estimate the productivity of a given school in a given year with its school-year fixed effect, \( q_{st} \), from the following regression, which is estimated across all fourth and fifth grade students from 1997 to 2011:

\[
y_{igst} = f_g(y_{ig-1's't-1}) + x'_{igst}\beta + q_{st} + \epsilon_{igst}. \tag{1.1}
\]

The standardized (at the grade-year level) math test score of student \( i \), who is in grade \( g \) at school \( s \) in time \( t \), is regressed on a school-year fixed effect along with the following controls: grade-specific cubic functions of the student’s prior math and reading scores; indicators for student grade repetition, ethnicity, gender, disability status, and limited English proficiency status; classroom averages of prior test scores and these demographic variables; and school-grade-year averages of prior test scores and demographic variables. The value-added productivity measures, \( q_{st} \), are estimates of the average test score gain students experience in year \( t \) as a result of being in school \( s \), while holding constant all other observable determinants of test scores.

If students were randomly assigned to schools, one could identify the causal effects of schools on student outcomes by comparing average student test scores across schools. The controls in equation (1.1) attempt to account for non-random student sorting across schools, implying that school value-added measures are biased when students differ across schools along unobservable characteristics that systematically affect test scores; in this case, the effects of these characteris-

\[13\]While third grade students can be accurately matched to their teachers, the data do not contain prior-year test scores (the most important control variable) for these students after 2005. I therefore drop third grade students and teachers from the analysis to ensure consistency throughout the sample from 1997 to 2011.
tics load onto the school-year fixed effects, confounding the interpretation of school value-added as the causal effect of attending a given school. Seemingly more productive schools, for example, may have students with parents who are more involved in their children’s education. When additional parental involvement leads to systematically higher test scores gains than the average gain captured by the control variables in equation (1.1), school-year effects capture both schools’ true contributions to test scores and the effects of parental involvement.

There exists relatively little work assessing whether school value-added measures are unbiased predictors of schools’ effects on student test scores. Deming (2014) and Angrist et al. (2015) both use admissions lotteries to assess whether value-added models produce unbiased estimates of school effects. Lotteries provide a unique opportunity to evaluate the contribution of each school to the test scores of students who were randomly given entry, thereby avoiding some of the selection issues in standard value-added models. As a test for unbiasedness of school value-added measures, both studies assess whether value-added estimates of school effects predict the estimates of school effects obtained using lottery admissions. Deming (2014) finds that value-added measures are unbiased estimates of school quality, while Angrist et al. (2015) use a more general test to detect some bias in conventional value-added measures.

As I am interested in measuring the change in school-level productivity that occurs as a result of competition, I explore how school value-added changes with competitive pressure. Neilson (2013) follows a similar approach when assessing how school quality responds to increased competitive pressure in Chile, Bau (2015) does the same using Pakistani data, and Dinerstein and Smith (2014) follow a similar approach when exploring the effects of a school finance reform on school productivity in New York. The focus on changes in school value-added implies that my results are biased when (a) school value-added measures partially reflect the test score effects of omitted student characteristics and (b) these characteristics change across schools after competition starts. I outline how I address such concerns in the discussion of the identification strategies below.
1.3.2 Measuring Teacher Productivity

To decompose school-level effects of competition into the effects of individual teachers, I match fourth and fifth grade students and teachers and estimate the following value-added regression with teacher-year effects:

\[ y_{ijgst} = h_g(y_{ijg-1's-1}) + x'_{ijgst} \gamma + q_{jt} + \epsilon_{ijgst}. \]  

Equation (1.2) estimates the value-added of each teacher \( j \) in year \( t \) as the teacher-year fixed effect \( q_{jt} \). The control variables in equation (1.2) include grade-specific cubic functions of students’ prior math and reading scores and indicators for student grade repetition, ethnicity, gender, disability status, and limited English proficiency status. Similar to school-level value-added, teacher value-added estimates represent the average contribution of teacher \( j \) to her students’ test scores.

In contrast to school value-added measures, there exists much empirical work assessing whether teacher value-added measures indeed capture causal effects on student outcomes or whether they are biased by student sorting across teachers. While there still exists some debate (Rothstein, 2010; Rothstein, 2014), much of the literature argues that teacher value-added measures are unbiased predictors of teacher performance and predict important short- and long-run effects of teachers on student outcomes (Kane and Staiger, 2008; Kane et al., 2013; Kane and Staiger, 2014; Chetty et al., 2014a; Chetty et al., 2014b; Bacher-Hicks et al., 2014). Most prior work treats teacher productivity as a relatively stable object, averaging teacher performance across many years to estimate teacher fixed effects. I use teacher-year fixed effects to measure a given teacher’s productivity in each year, which allows me to assess how teacher productivity changes when a charter school locates nearby. Previous studies have used teacher-year effect estimates to assess the properties of value-added estimators and to better understand the determinants of annual changes in teacher productivity (Aaronson et al., 2007; McCaffery et al., 2009; Goldhaber and Hansen, 2013).
1.3.3 Estimating the Effects of Competition on Productivity

Defining Competition

Researchers have used several methods to estimate the effects of charter school competition on TPS students,\(^{14}\) with studies generally differing by how they measure charter school competition and whether they use school- or student-level data to estimate its effects.

Hoxby (2003) uses school-level data from Michigan and Arizona, defining charter competition using minimum market penetration within districts and finding a significant association with school-level test score improvement and charter school penetration. Bettinger (2005) uses Michigan school-level data and defines the number of charter schools located within a 5-mile radius of a TPS as a measure of competitive pressure. Since state universities are authorizers of charter schools, the author instruments for the number of charter schools located near a TPS with the distance between the TPS and a state university, finding no significant effect on TPS test scores. Holmes et al. (2003) use distance from the nearest charter school and school-level data from North Carolina, showing that nearby charter school presence is associated with higher school-level test score improvements.

Estimating the effects of charter school competition using school-level data can lead to biased estimates when student composition changes after competition starts. Several researchers have recently turned to student-level data, employing student fixed effects to avoid confounding competitive effects with changes in unobserved student characteristics. These strategies identify the effects of charter school competition by comparing test score gains of the same TPS students before and after their TPSs face charter competition.

Bifulco and Ladd (2006) revisit competition in North Carolina using student-level data on third to eighth grade students from 1996 to 2002. Defining charter competition using distance to a charter school and the number of charter schools within various radii of a TPS, the authors find no effects on student test scores. Using student-level data from Florida, Sass (2006) measures competition using nearby charter school presence and market penetration, finding modest competitive effects, while Booker \textit{et al.} (2008) find similar results in Texas. Extending

\(^{14}\)See Betts (2009) for a full discussion of these methods.
the North Carolina data to 2005, Jinnai (2014) notes that charter schools exhibit the most competitive pressure when they offer grades that overlap with TPSs. Accounting for grade span overlap, he finds significant positive effects of charter school presence on TPS students.\footnote{Jackson (2012) analyzes the effects of charter presence on TPS teacher labor markets and also notes the importance of accounting for grade span overlap.}

In this chapter, I follow most prior work and rely on geographic proximity between charter schools and TPSs to define competition. My identification strategy (described below) involves assessing how the estimated productivity of a TPS changes when a charter school serving an overlapping grade with the TPS (from third to fifth grade) locates nearby. Using geographic distance to define competition implies that a TPS can enter competition in four ways: (i) a new charter school opens within a given distance of the TPS; (ii) a pre-existing charter school changes buildings, locating near the TPS; (iii) a pre-existing charter school expands its grade span to overlap with the TPS’s; or (iv) the TPS relocates near a pre-existing charter school.

To keep a tractable definition of competition, I focus on productivity changes that occur after sharp changes in competitive pressure, estimating competitive effects after a charter school opening event while excluding the other causes of competition. TPS responses to these other causes of competition may be more dynamic, beginning before a charter school option is available to students. For example, a TPS may feel competitive pressure before a charter school expands its grade span to overlap with the TPS’s third, fourth, or fifth grade.

To arrive at a variable capturing whether or not a TPS faces post-opening competition, I define the binary variable $CH_{st}^r$ as equal to one when there is a charter school serving an overlapping grade (from grades three to five) within $r$ miles of TPS $s$ in time $t$. Let the binary variable $PO_{st}^r$ equal one when a charter school serving an overlapping grade opened within $r$ miles of $s$ in some prior year $t' \leq t$ and there was previously no such charter school there. Using these two variables, I concisely define post-opening competition for TPS $s$ in time $t$ as

$$POC_{st}^r = 1(CH_{st}^r = 1 \text{ and } PO_{st}^r = 1),$$  \hspace{1cm} (1.3)
an overlapping grade within \( r \) miles and (ii) the first presence of such a charter school within \( r \) miles was caused by an opening event. Having defined competition between TPSs and charter schools, I now discuss how I estimate the effects of competition on TPS and teacher productivity while addressing several threats to identification.

**Estimating the Effects of Competition on School Productivity**

I estimate the effects of charter school competition on TPS productivity using the following difference-in-differences regression:

\[
\hat{\theta}_{st} = \text{POC}_{st} \rho^r + \mathbf{x}'_{st} \gamma + \lambda_t + \psi_s + \epsilon_{st},
\]

where the dependent variable is the estimated school-year fixed effect from equation (1.1). The main parameter of interest is \( \rho^r \), which measures the effect of charter school competition on TPS productivity. As described above, \( \text{POC}_{st} \) is a binary variable denoting post-opening charter school competition within a \( r \)-mile radius of TPS \( s \). Although most studies use radii ranging from 1 to 10 miles, the appropriate choice of \( r \) depends on urban density and the size of typical school attendance zones in a given area (Imberman, 2011). Bifulco and Ladd (2006) find that approximately 90 percent of all student transfers between TPSs and charter schools in North Carolina occur across schools within a 10-mile radius of each other, suggesting that competitive effects could operate even when a charter school is 10 miles away. I explore the sensitivity of the results to several different values of \( r \).

The control vector \( \mathbf{x}_{st} \) contains school-year averages of students’ prior-year math and reading scores, and school-year percentages of minority students, disabled students, limited English proficient students, free-lunch-eligible students, and male students. To ensure that the estimate of charter school competition is not confounded by time trends in state-wide productivity, I also control for year fixed effects \( \lambda_t \).

Since charter school entry is not random, even after including these controls, a comparison between schools that face charter school competition and those that do not can result in biased estimates of competitive effects. The bias can be positive or negative, depending on
the correlation between charter school entry decisions and TPS quality. To account for such correlation, I include school fixed effects, $\psi_s$, in equation (1.4), implying that $\rho^c$ is identified by comparing the quality of the same school before and after it faces competition with a charter school. This strategy circumvents any threats to identification that result from charter school location decisions being correlated with fixed determinants of TPS quality.

There are two remaining threats to the identification of competitive effects. First, as mentioned in Section 1.2, charter schools have strong incentives to locate in areas where demand for the type of education they provide is sufficiently high. Charter school founders are likely aware of changing neighborhood trends and may make location decisions based on changes in TPS characteristics that are correlated with changes in TPS productivity. As neighborhoods slowly gentrify, for instance, parents may become more involved in local education and pressure the local TPS to gradually become more productive. At the same time, the growing demand for education quality may make a charter school more likely to enter the market in order to meet that demand. In this case, not accounting for the improvement trend in TPS productivity prior to charter entry would result in upward biased estimates of competitive effects.

I rule out such sources of bias in two ways. First, I take advantage of the longitudinal nature of the data and conduct event-studies in which I plot trends in TPS productivity before and after a charter school opens nearby. As shown below, the event-studies show no evidence of a systematic trend in productivity prior to a charter school opening. Second, I assess whether the results are sensitive to restricting the analysis sample to only TPSs that eventually face charter school competition. The difference-in-differences estimator identifies the effect of competition by comparing the change in productivity of TPSs that had a charter school open nearby to the change in productivity of TPSs that did not experience a charter school opening throughout the same time period. If charter schools only open near TPSs with specific trends in productivity, using TPSs that never face competition to form part of the counterfactual trend in the difference-in-differences estimator may result in a biased estimate. Restricting the sample to only TPSs that eventually face competition yields nearly identical results, which is consistent with charters not systematically entering areas based on particular trends in TPS productivity.

The second threat to identification is charter school entry causing students to sort across
schools. For example, if poorly-performing students exit a TPS in response to charter school entry, an improvement in TPS productivity could reflect a more able student body rather than a productivity response. In contrast, if charter school presence results in high-achieving students departing, one could erroneously conclude that charter schools harm students in TPSs. The estimation of school-year quality controls for rich polynomials in students’ prior scores and equation (1.4) controls for school-year averages of prior scores and student characteristics. It is still possible, however, that school-year value-added measures partially capture the effects of unobserved test score determinants. If these unobserved characteristics change in a systematic way when students re-sort after a charter school opens, my identification strategy will capture a student composition effect in addition to a TPS productivity effect.

I use event-studies to show that there are minimal changes in observable student characteristics at TPSs when charter schools open nearby. While I find no evidence of systematic changes in important observable determinants of test scores, such as prior scores, charter school entry may cause students to sort along unobserved test score determinants. I directly assess the robustness of the results to changes in unobservable student characteristics by following previous work and estimating the effects of competition on student test scores using student-level regressions with student-school fixed (spell) effects (Bifulco and Ladd, 2006; Sass, 2006; Booker et al., 2008; Imberman, 2011). The student-level regressions yield estimated effects very similar to those from the main analysis.

**Estimating the Effects of Competition on Teacher Productivity**

In terms of its teaching staff, a TPS can become more productive in two ways: it can either hire new, more productive teachers or it can work with its current staff to make them more productive. If TPSs hire new teachers or reassign teachers across tested and non-tested grades, one would not expect to see any effect on the productivity of individual teachers when TPSs become better. In contrast, if existing teachers improve, one should be able to detect within-teacher improvements in performance.

I estimate the effects of working in a TPS facing competition on individual teachers using
the following teacher-level difference-in-differences regression:

\[
\hat{q}_{jt} = POC_{rt} \beta + x_{jst}' \gamma + \lambda_t + \psi_j + \epsilon_{jst},
\]

(1.5)

where the dependent variable is the teacher-year fixed effect from equation (1.2). The main parameter of interest is \( \beta \), which measures the effect on teacher productivity of working in a TPS facing charter school competition. The control vector \( x_{jst} \) contains grade fixed effects, controls for teacher experience,\(^{16}\) classroom, school-grade-year, and school-year averages of students’ prior-year math and reading scores, and percentages of minority students, disabled students, limited English proficient students, free-lunch-eligible students,\(^{17}\) and male students. As before, year fixed effects \( \lambda_t \) capture state-wide trends in teacher performance. The inclusion of teacher fixed effects \( \psi_j \) in equation (1.5) ensures that the effect of working in a TPS facing competition is identified by within-teacher changes in value-added. The identification strategy is robust to teachers selecting into TPSs facing competition based on fixed characteristics or ability, as these traits are captured by the fixed effects.

There are two remaining threats to identification. First, teachers may sort into competitive environments based on projected changes in productivity. For instance, a teacher may be steadily improving and choose to stay in or move to a TPS facing competition in order to work in a challenging environment. In cases like these, the change in productivity starts to occur prior to competition and not accounting for it results in biased estimates of competitive effects. I address these concerns with event-studies, showing productivity trends several years before and after competition. The event-studies show no systematic trends prior to a teacher working in a competitive TPS.

The second threat to identification relates to confounding competitive effects with teachers switching schools and improving performance as result of finding a better match with their new schools. Jackson (2013) shows that teacher performance significantly improves after switching schools, which is consistent with teachers seeking improvements in match quality. If teachers

\(^{16}\) I include indicator variables for zero, one, two, three, four, and five years of experience. The omitted category is six or more years of teaching experience.

\(^{17}\) Free-lunch status is only available at the school-year-level (not at the student-level) in 1997 and 1998. To be able to use the full sample in the regression, I only include this variable at the school-year level.
switch schools prior to beginning to work in a competitive TPS, not accounting for match quality upgrading could lead to upward biased estimates of competitive effects. I control for match quality by augmenting equation (1.5) to account for time-since-switching schools and by restricting the sample to teachers who never switch. Both methods yield nearly identical results to the main analysis.

1.4 The School-Level Effects of Competition

In this section, I present the estimated effects of charter school competition on TPS productivity. Table 1.4 shows the effects on TPS productivity from different regressions based on the distance used to define competition. Panel (a) reports the estimated effects of facing charter school competition on school math value-added, while Panel (b) shows the effects of observing a TPS in the period after a charter school opened but when one is not necessarily still nearby, thus allowing any productivity effects to persist even after a potential charter school exit. Panels (c) and (d) show the corresponding effects on school reading value-added.

For math, statistically significant competitive effects begin to emerge at the 4-mile mark and persist for radii as large as 10 miles. These estimates suggest that a TPS raises student math scores by 0.025 standard deviations more when it faces competition from a charter school, an estimate well within the range found by previous work using student-level regressions. While these effects seem modest at the student-level, they represent a 0.12 standard deviation improvement in school-level math productivity. Panels (c) and (d) show the estimated effects on TPS reading productivity. In contrast to math scores, I estimate no significant effect of charter school competition on TPS reading value-added, which is consistent with schools and teachers having greater effects on student math performance, a common finding in the literature.\footnote{See, for example, Sass (2006).}

It is somewhat counter-intuitive that competitive effects do not emerge at distances less than 4 miles away from TPSs. Reassuringly, Figure 1.1 shows that the effects of competition at smaller radii are statistically indistinguishable from those found at a radius of 5 miles. There are relatively few TPSs that face charter school competition at distances closer than 4 miles,
leading to noisy estimates of competition when using these radii to define competitive pressure. To improve the precision of my estimates, I use a radius of 5 miles to define competition throughout the remainder of the chapter.

1.4.1 Robustness Checks

In this subsection, I address the main threats to identification of competitive effects: charter schools locating in areas based on trends in TPS productivity and student sorting after a charter school opens nearby.

Charter Entry Based on Trends

I explore time trends in school quality before and after charter school competition by using the rich longitudinal data to estimate the following event-study regression:

\[ \hat{q}_{st} = 1(t - \tau^*_s \leq -4) \rho_{-4} + \sum_{k=-3}^{2} 1(t - \tau^*_s = k) \rho_k + 1(t - \tau^*_s \geq 3) \rho_3 + x'_{st} \gamma + \lambda_t + \psi_s + \epsilon_{st}, \quad (1.6) \]

where \( \tau^*_s \) denotes the year TPS \( s \) first started competition with a charter school. For \(-3 \leq k \leq 2\), each \( \rho_k \) coefficient measures average TPS productivity \( k \) years from the start of competition, while \( \rho_{-4} \) and \( \rho_3 \) measure average TPS productivity in 4 or more years before and 3 or more years after the start of competition, respectively. Figure 1.2 plots the \( \rho_k \) estimates along with the associated 90 percent confidence intervals. Panels (a) and (b) show the effects on TPS math value-added while panels (c) and (d) show the effects on reading value-added.

Panel (a) shows no evidence of a pre-existing trend in TPS math productivity prior to a charter school opening nearby and clear evidence of an improvement in productivity after competition starts. In panel (b), I restrict the sample to TPS that ever faced competition to assess whether these schools exhibit differential productivity trends. The estimated effects before and after competition remain virtually unchanged. Consistent with the results in panels (c) and (d) of Table 1.4, panels (c) and (d) of Figure 1.2 show no significant improvement in TPS reading productivity after a charter school opens nearby. Reassuringly, there is also no evidence of a pre-existing trend prior to opening events in both the full and ever-treated samples.
Student Sorting and Enrollment Changes

In panel (a) of Figure 1.3, I plot trends in the fraction of third to fifth grade students in a TPS who were observed in the TPS in the prior year but attend a charter school in the current year. The reported coefficients are regression-adjusted for year and TPS fixed effects. There is a clear jump in the fraction of students transferring into the charter school system when a charter school opens nearby a TPS, with the transfer rate jumping by 1.29 percentage points. Although this is a small fraction in absolute terms, it is more than 2.5 times the magnitude of the mean transfer rate across all TPS-year observations (0.005 percent).

Panel (b) shows that third to fifth grade enrollment is growing slightly prior to charter school entry and then stabilizes after, while panel (c) shows a similar pattern for the number of full-time equivalent teachers employed by TPSs. Since both enrollment and the number of teachers first grows and then stabilizes, average class size remains relatively flat over time, as shown by panel (d). It is therefore unlikely that the productivity effects documented above reflect changes in class sizes at TPSs.

In Figure 1.4, I assess how observable student characteristics change at TPSs when a charter school locates nearby. The reported coefficients are again regression-adjusted only for year and TPS fixed effects. Given the relatively small (absolute) effects on enrollment, it is unsurprising that student characteristics do not change immediately after an opening event. Panels (a) and (b) of Figure 1.4 show no changes in average prior math and reading scores, while panel (c) shows no change in the fraction of students eligible for a free or reduced-price lunch. Panel (d) shows a steady decline in the fraction of students with college-educated parents, which is consistent with more educated parents being more likely to exercise school choice and leave the TPS system. Panels (e), (f), and (g) show no changes in the fractions of disabled, male, or limited English proficient students. Panel (h) shows a steady increase in the fraction of minority (non-white) students.

Overall, these analyses reveal minimal changes in observable student characteristics.\textsuperscript{19} The

\textsuperscript{19}Despite the small changes, when estimating productivity effects in equation (1.4), I control for all of these school-level characteristics except the fraction of students with college-educated parents. Since parental education is only available from 1997 to 2006, I drop it from the analyses to be able to use the entire sample until 2011. Restricted sample regressions including it as a control yield nearly identical results (available upon request).
only exceptions are the decline in the fraction of students with college-educated parents and the increase in the fraction of minority students. Note that, since parental education is positively correlated with academic achievement and minority status is negatively correlated, if unobserved characteristics change according to similar sorting patterns, such changes should lead to negative bias in the productivity effects documents above, making it harder to detect any effect.

To directly address concerns over student sorting along unobservable dimensions, I follow prior work and estimate student-level regressions with student-school (spell) fixed effects:

\[ y_{ist} = POC_{st}^r \rho + X_{ist}' \xi + \mu_{is} + \epsilon_{ist}. \]  

(1.7)

The control vector contains year and grade fixed effects along with classroom, school-grade-year, and school-year percentages of minority students, limited English proficient students, disabled students, and male students. The inclusion of student-school fixed effects \( \mu_{is} \) implies that the effect of charter school competition on TPS student test scores is identified by comparing achievement of the same students before and after their TPS faces competition, while eliminating any fixed student and school characteristics that are associated with attending a TPS that is close to a charter school.

Table 1.5 presents the results from variants of equation (1.7) using radii of 4, 5, and 6 miles to define competition. The specification in panel (a) does not use student-school fixed effects, instead relying on a rich set of student-level control variables, including grade-specific cubic functions of prior scores. The estimated effect of charter school competition is very similar to the effect found using school-level productivity measures. In panel (b), I drop the student-level controls and include student-school fixed effects. The estimates become much noisier but remain very similar in magnitude to the school-level productivity effect documented above. In panel (c), I change the dependent variable to math score gains, which again increases the standard errors but does not statistically change the point estimates.
1.5 The Teacher-Level Effects of Competition

Having shown clear evidence that TPSs become more productive after facing charter school competition, I now investigate whether they achieve these gains by hiring more productive teachers or by a given group of teachers becoming more productive. To start, Figure 1.5 shows event-study analyses for teacher turnover using a specification similar to equation (1.6). Panel (a) shows that the fraction of newly hired teachers at the average TPS falls significantly after a charter school opens nearby, dropping by an estimated 5 percentage points, or 24 percent of the mean. Coupled with the small (statistically insignificant) decline in the fraction of teachers leaving TPSs shown in panel (b), this evidence suggests that TPSs do not rely on changes in teaching staff to become more productive after entering competition with a charter school.

Indeed, Table 1.6 shows the results from estimating variants of equation (1.5) and provides clear evidence that school-level productivity improvements are in fact driven by within-teacher productivity improvements, as the same teachers perform better when working in a TPS facing competition. In column (1), I use the full sample of TPS teachers. Panel (a) shows that a given teacher has a 0.02 student-level standard deviations higher math value-added when working in a TPS facing competition from a charter school. In terms of the distribution of teacher productivity, this amounts to a given teacher being 0.09 standard deviations more productive when working in a competitive TPS. Panel (b) shows that the performance improvement persists even after teachers no longer work in such a school, as the effect of being observed after working in a competitive TPS is even higher at 0.13 teacher-level standard deviations. In column (2), I replace teacher fixed effects with teacher-school fixed effects, thereby identifying the effect of working in a competitive TPS primarily with those teachers who are observed in the same school before and after a charter school opening. Although the coefficients drop slightly, they remain very similar to those from column (1) and highly significant.

\[^{20}\text{Jackson (2012) finds a similar result, interpreting the reduction in new hires as the charter school competitor representing an additional employer with which the TPS must compete for teachers.}\]
1.5.1 Robustness Checks

In this subsection, I assess whether the results are partially driven by teachers sorting into competitive TPSs based on projected trends in productivity or match quality upgrading between teachers and schools.

Trends in Productivity

Figure 1.6 plots event studies at the teacher-level, showing trends in productivity several years before and after a teacher works in a TPS facing competition. I create these plots by estimating equation (1.5) while replacing the $POC_{ijt}^r$ variable with a full set of indicators for the time since the teacher began working in a competitive environment. Panel (a) shows a clear improvement in teacher math value-added after working in a TPS facing competition and no systematic trend in productivity beforehand. The absence of a clear trend prior to competition suggests it is unlikely that teachers sort into competitive environments based on anticipated changes in performance.

Match Quality Upgrading

It is possible, however, that teachers switch schools in an effort to find a better school match. If school switching times overlap significantly with the period around which teachers begin working in a TPS facing competition, the results may reflect productivity gains due to teacher-school match quality improvements rather than competitive pressure. Jackson (2013) shows compelling evidence of match quality improvement following teacher switching, using an event-study with ‘time-since-switching-schools’ indicators to show a clear and sustained jump in productivity in the year of switching.

To test whether match quality improvements drive my results, I first include a post-switch indicator variable and re-estimate equation (1.5). Column (3) of Table 1.6 shows that the results from this specification are nearly identical to the results from column (1). In columns (4) and (5), I further show the results are robust to estimating the effects of competition using either the sample of teachers who never switch schools or only the sample of teachers who do switch schools. In panel (b) of Figure 1.6, I include a full set of ‘time-since-switching-schools’
indicators in the regression and still estimate virtually the same event-study coefficients as in panel (a), while panel (c) produces similar results using only the sample of teachers who never switch schools. Although they are not reported, the estimates of the ‘time-since-switching-schools’ indicators from this regression are very similar to the results reported in Jackson (2013), implying that match quality upgrading is an important determinant of teacher productivity but one that operates independently of competitive pressure.

In sum, there is robust evidence that the aggregate school-level effect of charter school competition is driven by within-teacher productivity improvements.

1.6 Conclusion

In this chapter, I presented evidence indicating that TPSs respond to competition from charter schools in an environment where strong accountability incentives prevailed prior to competition. Using rich school administrative data, I showed that school-level productivity improvements are entirely driven by the existing teaching staff at TPSs becoming more productive. This is a new finding in both the literature examining the effects of school competition, where most previous work has argued that competitive pressure should lead to hiring more productive teachers instead of making the current staff better, and the literature concerned with measuring teacher productivity, where the degree of market competition has not been previously considered as determinant of teacher performance.

In the next chapter, I examine the relationship between the incentives created by increased school competition and those stemming from school-based accountability programs. Specifically, I investigate why and how competitive forces cause schools to improve performance, even in the presence of the high-powered incentives they already face. Understanding such interactions is relevant to policy, as the widespread presence of school-based accountability programs across public education systems in the United States guarantees that future competition based policies – whether they be private school vouchers or charter school expansion – will be enacted in environments where educators already face explicit performance incentives.
Table 1.1: Student Characteristics in Schools Serving Grades 3 to 5 from 1997 to 2011

<table>
<thead>
<tr>
<th></th>
<th>(1) All TPS Students</th>
<th>(2) Valid for VA TPS Students</th>
<th>(3) TPS Students Near Charters</th>
<th>(4) TPS Students Not Near Charters</th>
<th>(5) Charter School Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math Score</td>
<td>0</td>
<td>0.05</td>
<td>0.04</td>
<td>-0.01</td>
<td>-0.13</td>
</tr>
<tr>
<td>(1)</td>
<td>(0.98)</td>
<td>(1.03)</td>
<td>(0.98)</td>
<td>(1.02)</td>
<td></td>
</tr>
<tr>
<td>Prior Math Score</td>
<td>0.01</td>
<td>0.04</td>
<td>0.04</td>
<td>-0.01</td>
<td>-0.09</td>
</tr>
<tr>
<td>(1)</td>
<td>(0.98)</td>
<td>(1.04)</td>
<td>(0.98)</td>
<td>(1.03)</td>
<td></td>
</tr>
<tr>
<td>Reading Score</td>
<td>0</td>
<td>0.04</td>
<td>0.03</td>
<td>-0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>(1)</td>
<td>(0.98)</td>
<td>(1.03)</td>
<td>(0.99)</td>
<td>(1.03)</td>
<td></td>
</tr>
<tr>
<td>Prior Reading Score</td>
<td>0</td>
<td>0.04</td>
<td>0.03</td>
<td>-0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>(1)</td>
<td>(0.98)</td>
<td>(1.03)</td>
<td>(0.99)</td>
<td>(1.03)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.51</td>
<td>0.5</td>
<td>0.51</td>
<td>0.51</td>
<td>0.5</td>
</tr>
<tr>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.5)</td>
</tr>
<tr>
<td>Limited English Proficient Student</td>
<td>0.05</td>
<td>0.03</td>
<td>0.07</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>(0.21)</td>
<td>(0.17)</td>
<td>(0.25)</td>
<td>(0.2)</td>
<td>(0.11)</td>
<td></td>
</tr>
<tr>
<td>Student with College-Educated Parents†</td>
<td>0.25</td>
<td>0.26</td>
<td>0.34</td>
<td>0.22</td>
<td>0.42</td>
</tr>
<tr>
<td>(0.43)</td>
<td>(0.44)</td>
<td>(0.47)</td>
<td>(0.42)</td>
<td>(0.49)</td>
<td></td>
</tr>
<tr>
<td>Free or Reduced-Price Lunch Eligible Student‡†</td>
<td>0.48</td>
<td>0.44</td>
<td>0.47</td>
<td>0.48</td>
<td>0.24</td>
</tr>
<tr>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.43)</td>
<td></td>
</tr>
<tr>
<td>Disabled Student</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>(0.22)</td>
<td>(0.21)</td>
<td>(0.22)</td>
<td>(0.22)</td>
<td>(0.22)</td>
<td></td>
</tr>
<tr>
<td>Racial Minority Student</td>
<td>0.42</td>
<td>0.4</td>
<td>0.53</td>
<td>0.38</td>
<td>0.41</td>
</tr>
<tr>
<td>(0.49)</td>
<td>(0.49)</td>
<td>(0.5)</td>
<td>(0.49)</td>
<td>(0.49)</td>
<td></td>
</tr>
<tr>
<td>Student Repeating Current Grade</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>(0.12)</td>
<td>(0.12)</td>
<td>(0.12)</td>
<td>(0.12)</td>
<td>(0.14)</td>
<td></td>
</tr>
<tr>
<td>Student-Year Observations</td>
<td>4,692,842</td>
<td>2,499,596</td>
<td>1,349,599</td>
<td>3,343,243</td>
<td>88,972</td>
</tr>
<tr>
<td>Student Observations</td>
<td>1,937,208</td>
<td>1,347,827</td>
<td>640,436</td>
<td>1,478,800</td>
<td>47,753</td>
</tr>
</tbody>
</table>

Notes: Summary statistics are calculated over all third to fifth grade students from 1997 to 2011. Standard deviations are in parentheses in the second row of each cell. A student attends a TPS that is near a charter school if the TPS is located within 5 miles of a charter school serving a grade from three to five that overlaps with the TPS. † Parental Education is only available from 1997 to 2006. ‡‡ Free-lunch eligibility is only available from 1999 to 2011.
Table 1.2: Schools Serving Grades 3 to 5 from 1997 to 2011

<table>
<thead>
<tr>
<th>Restriction:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All TPSs</td>
<td>TPSs w/n 1 Mile of Charter</td>
<td>TPSs w/n 5 Miles of Charter</td>
<td>TPSs w/n 10 Miles of Charter</td>
<td>TPSs &gt; 10 Miles of Charter</td>
<td>Charter Schools</td>
</tr>
<tr>
<td>Third to Fifth Grade</td>
<td>258.24</td>
<td>226.72</td>
<td>269.21</td>
<td>271.26</td>
<td>245.38</td>
<td>118.15</td>
</tr>
<tr>
<td>Student Enrollment</td>
<td>(107.07)</td>
<td>(85.57)</td>
<td>(102.48)</td>
<td>(105.44)</td>
<td>(107.13)</td>
<td>(73.11)</td>
</tr>
<tr>
<td>Highest Grade Served</td>
<td>5.44</td>
<td>5.49</td>
<td>5.34</td>
<td>5.32</td>
<td>5.55</td>
<td>7.78</td>
</tr>
<tr>
<td>(1.44)</td>
<td>(1.86)</td>
<td>(1.45)</td>
<td>(1.32)</td>
<td>(1.52)</td>
<td>(2.26)</td>
<td></td>
</tr>
<tr>
<td>Lowest Grade Served</td>
<td>0.29</td>
<td>0.21</td>
<td>0.21</td>
<td>0.21</td>
<td>0.37</td>
<td>0.38</td>
</tr>
<tr>
<td>(1.05)</td>
<td>(0.88)</td>
<td>(0.87)</td>
<td>(0.88)</td>
<td>(1.17)</td>
<td>(1.22)</td>
<td></td>
</tr>
<tr>
<td>Located in Urban Area</td>
<td>0.56</td>
<td>0.89</td>
<td>0.81</td>
<td>0.7</td>
<td>0.45</td>
<td>0.68</td>
</tr>
<tr>
<td>(0.5)</td>
<td>(0.32)</td>
<td>(0.39)</td>
<td>(0.46)</td>
<td>(0.5)</td>
<td>(0.47)</td>
<td></td>
</tr>
<tr>
<td>Average Student Math Score</td>
<td>-0.02</td>
<td>-0.15</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.06</td>
<td>-0.24</td>
</tr>
<tr>
<td>(0.37)</td>
<td>(0.45)</td>
<td>(0.42)</td>
<td>(0.41)</td>
<td>(0.32)</td>
<td>(0.55)</td>
<td></td>
</tr>
<tr>
<td>Math Value-Added</td>
<td>0</td>
<td>0.01</td>
<td>0.04</td>
<td>0.03</td>
<td>-0.04</td>
<td>-0.21</td>
</tr>
<tr>
<td>(0.2)</td>
<td>(0.21)</td>
<td>(0.2)</td>
<td>(0.2)</td>
<td>(0.2)</td>
<td>(0.32)</td>
<td></td>
</tr>
<tr>
<td>Average Student Reading Score</td>
<td>-0.03</td>
<td>-0.15</td>
<td>-0.02</td>
<td>0.01</td>
<td>-0.06</td>
<td>-0.07</td>
</tr>
<tr>
<td>(0.34)</td>
<td>(0.45)</td>
<td>(0.41)</td>
<td>(0.39)</td>
<td>(0.29)</td>
<td>(0.52)</td>
<td></td>
</tr>
<tr>
<td>Reading Value-Added</td>
<td>-0.01</td>
<td>0</td>
<td>0.03</td>
<td>0.02</td>
<td>-0.04</td>
<td>-0.05</td>
</tr>
<tr>
<td>(0.15)</td>
<td>(0.18)</td>
<td>(0.16)</td>
<td>(0.16)</td>
<td>(0.14)</td>
<td>(0.24)</td>
<td></td>
</tr>
<tr>
<td>Fraction of Students with College-Educated Parents†</td>
<td>0.23</td>
<td>0.27</td>
<td>0.32</td>
<td>0.3</td>
<td>0.18</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.25)</td>
<td>(0.24)</td>
<td>(0.23)</td>
<td>(0.14)</td>
<td>(0.26)</td>
</tr>
<tr>
<td>Fraction of Free-Lunch Eligible Students††</td>
<td>0.49</td>
<td>0.6</td>
<td>0.51</td>
<td>0.48</td>
<td>0.5</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>(0.23)</td>
<td>(0.29)</td>
<td>(0.26)</td>
<td>(0.25)</td>
<td>(0.22)</td>
<td>(0.33)</td>
</tr>
<tr>
<td>Fraction of Minority Students</td>
<td>0.44</td>
<td>0.67</td>
<td>0.56</td>
<td>0.49</td>
<td>0.38</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>(0.29)</td>
<td>(0.3)</td>
<td>(0.28)</td>
<td>(0.29)</td>
<td>(0.27)</td>
<td>(0.36)</td>
</tr>
<tr>
<td>Number of Charters with Overlapping Grade Nearby</td>
<td>-</td>
<td>1.1</td>
<td>1.69</td>
<td>2.38</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.35)</td>
<td>(1.18)</td>
<td>(1.91)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Passed ABCs</td>
<td>0.78</td>
<td>0.72</td>
<td>0.8</td>
<td>0.8</td>
<td>0.75</td>
<td>0.63</td>
</tr>
<tr>
<td>(0.42)</td>
<td>(0.45)</td>
<td>(0.4)</td>
<td>(0.4)</td>
<td>(0.43)</td>
<td>(0.48)</td>
<td></td>
</tr>
<tr>
<td>Passed NCLB</td>
<td>0.6</td>
<td>0.53</td>
<td>0.55</td>
<td>0.59</td>
<td>0.61</td>
<td>0.65</td>
</tr>
<tr>
<td>(0.49)</td>
<td>(0.5)</td>
<td>(0.49)</td>
<td>(0.49)</td>
<td>(0.49)</td>
<td>(0.48)</td>
<td></td>
</tr>
<tr>
<td>School-Year Observations</td>
<td>19,853</td>
<td>610</td>
<td>5,371</td>
<td>9,258</td>
<td>10,595</td>
<td>927</td>
</tr>
</tbody>
</table>

Notes: Summary statistics are calculated with all charter schools and TPSs serving at least one grade from three to five. The sample runs from 1997 to 2011 and the unit of observation is a school-year. Standard deviations are in parentheses in the second row of each cell. A TPS is within a given distance of a charter school only if that charter school serves a grade from three to five that overlaps with the TPS. † Parental Education is only available from 1997 to 2006. †† In contrast to the student-level data, the fraction of students who are eligible for a free or reduced-price lunch is available at the school-level throughout the entire sample from 1997 to 2011.
Table 1.3: Teacher Characteristics in Schools Serving Grades 3 to 5 from 1997 to 2011

<table>
<thead>
<tr>
<th></th>
<th>(1) All TPS Teachers</th>
<th>(2) VA-Eligible Teachers</th>
<th>(3) Teachers in TPSs Near Charter</th>
<th>(4) Teachers not in TPSs Near Charter</th>
<th>(5) Charter Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of Experience</td>
<td>12.33 (9.5)</td>
<td>12.01 (9.67)</td>
<td>11.86 (9.48)</td>
<td>12.53 (9.51)</td>
<td>8.49 (7.86)</td>
</tr>
<tr>
<td>Female</td>
<td>0.92 (0.27)</td>
<td>0.91 (0.29)</td>
<td>0.92 (0.27)</td>
<td>0.92 (0.27)</td>
<td>0.82 (0.38)</td>
</tr>
<tr>
<td>Racial Minority</td>
<td>0.16 (0.37)</td>
<td>0.15 (0.36)</td>
<td>0.21 (0.4)</td>
<td>0.15 (0.36)</td>
<td>0.24 (0.42)</td>
</tr>
<tr>
<td>Has Teacher-Year Value-Added (VA)</td>
<td>0.13 (0.34)</td>
<td>0 (0)</td>
<td>0.13 (0.33)</td>
<td>(0.13 (0.34)</td>
<td>-</td>
</tr>
<tr>
<td>Teacher-Year Math VA</td>
<td>0 (0.24)</td>
<td>0 (0.24)</td>
<td>0.03 (0.23)</td>
<td>-0.01 (0.25)</td>
<td>-</td>
</tr>
<tr>
<td>Teacher-Year Reading VA</td>
<td>0 (0.18)</td>
<td>0 (0.18)</td>
<td>0.02 (0.18)</td>
<td>-0.01 (0.18)</td>
<td>-</td>
</tr>
<tr>
<td>Teacher-School-Year Obs.</td>
<td>707,734</td>
<td>92,468</td>
<td>209,378</td>
<td>498,356</td>
<td>21,374</td>
</tr>
<tr>
<td>Teacher Obs.</td>
<td>136,865</td>
<td>26,133</td>
<td>54,807</td>
<td>114,499</td>
<td>7,842</td>
</tr>
</tbody>
</table>

Notes: Summary statistics are calculated over all teacher-school-year observations in schools serving grades three to five from 1997 to 2011. Standard deviations are in parentheses in the second row of each cell. A teacher works in a TPS that is near a charter school if the TPS is located within 5 miles of a charter school that serves a grade from three to five that overlaps with the TPS.
Table 1.4: Effects of Competition on School Math and Reading Value-Added (VA)

<table>
<thead>
<tr>
<th>Mile Radius:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>POC</td>
<td>-0.0081</td>
<td>0.0152</td>
<td>0.0091</td>
<td>0.0197***</td>
<td>0.0243***</td>
<td>0.0273***</td>
<td>0.0283***</td>
<td>0.0268***</td>
<td>0.0268***</td>
<td>0.0259***</td>
</tr>
<tr>
<td></td>
<td>(0.0218)</td>
<td>(0.0117)</td>
<td>(0.0084)</td>
<td>(0.0075)</td>
<td>(0.0068)</td>
<td>(0.0069)</td>
<td>(0.0070)</td>
<td>(0.0067)</td>
<td>(0.0069)</td>
<td>(0.0068)</td>
</tr>
<tr>
<td>PO</td>
<td>0.0043</td>
<td>0.0161</td>
<td>0.0112</td>
<td>0.0156*</td>
<td>0.0252***</td>
<td>0.0265***</td>
<td>0.0263***</td>
<td>0.0233***</td>
<td>0.0270***</td>
<td>0.0225***</td>
</tr>
<tr>
<td></td>
<td>(0.0272)</td>
<td>(0.0141)</td>
<td>(0.0106)</td>
<td>(0.0091)</td>
<td>(0.0084)</td>
<td>(0.0083)</td>
<td>(0.0082)</td>
<td>(0.0079)</td>
<td>(0.0081)</td>
<td>(0.0079)</td>
</tr>
<tr>
<td>POC</td>
<td>-0.0127</td>
<td>0.0004</td>
<td>-0.0058</td>
<td>0.0045</td>
<td>0.0039</td>
<td>0.0067</td>
<td>0.0055</td>
<td>0.0076*</td>
<td>0.0091**</td>
<td>0.0086**</td>
</tr>
<tr>
<td></td>
<td>(0.0132)</td>
<td>(0.0082)</td>
<td>(0.0061)</td>
<td>(0.0054)</td>
<td>(0.0050)</td>
<td>(0.0045)</td>
<td>(0.0044)</td>
<td>(0.0042)</td>
<td>(0.0041)</td>
<td>(0.0040)</td>
</tr>
<tr>
<td>PO</td>
<td>-0.0117</td>
<td>-0.0016</td>
<td>-0.0066</td>
<td>-0.0022</td>
<td>0.0007</td>
<td>0.0020</td>
<td>0.0017</td>
<td>0.0035</td>
<td>0.0038</td>
<td>0.0033</td>
</tr>
<tr>
<td></td>
<td>(0.0135)</td>
<td>(0.0092)</td>
<td>(0.0070)</td>
<td>(0.0060)</td>
<td>(0.0054)</td>
<td>(0.0050)</td>
<td>(0.0049)</td>
<td>(0.0048)</td>
<td>(0.0048)</td>
<td>(0.0047)</td>
</tr>
<tr>
<td>Observations</td>
<td>16,221</td>
<td>16,221</td>
<td>16,221</td>
<td>16,221</td>
<td>16,221</td>
<td>16,221</td>
<td>16,221</td>
<td>16,221</td>
<td>16,221</td>
<td>16,221</td>
</tr>
</tbody>
</table>

Notes: The dependent variable in each column of panels (a) and (b) is school-year math value-added. The dependent variable in each column of panels (c) and (d) is school-year reading value-added. Each cell represents an estimate from a separate regression and each regression is estimated using school-level data from 1997 to 2011 for school-year observations with an estimated value-added across grades four or five in math and reading. All regressions include year and school fixed effects and the following control variables: school-year averages of student prior math and reading scores and school-year percentages of minority students, disabled students, limited English proficient students, male students, and free-lunch eligible students. Each cell in panels (a) and (c) reports the effect of post-opening charter school competition within the distance given under the respective column number. Each cell in panels (b) and (d) reports the effect of being in the post-opening period (when there is not necessarily a charter school still operating nearby) within the distance given under the respective column number. Standard errors clustered at the school level appear in parentheses. *** denotes significance at the 1% level; ** denotes significance at the 5% level; * denotes significance at the 10% level.
Chapter 1. School Competition, Accountability, and Teacher Productivity

Table 1.5: Student-Level Regression Results

<table>
<thead>
<tr>
<th>Mile Radius:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤4m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤5m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤6m</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a): Value-Added Regression

<table>
<thead>
<tr>
<th>Post-Opening Competition</th>
<th>0.0252****</th>
<th>0.0318****</th>
<th>0.0320****</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.0081)</td>
<td>(0.0074)</td>
<td>(0.0074)</td>
</tr>
</tbody>
</table>

Student-School Fixed Effects: N  N  N

(b): Level Score as Dependent Variable

<table>
<thead>
<tr>
<th>Post-Opening Competition</th>
<th>0.0271**</th>
<th>0.0246*</th>
<th>0.0199</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.0137)</td>
<td>(0.0129)</td>
<td>(0.0125)</td>
</tr>
</tbody>
</table>

Student-School Fixed Effects: Y  Y  Y

(c): Gain Score as Dependent Variable

<table>
<thead>
<tr>
<th>Post-Opening Competition</th>
<th>0.0265</th>
<th>0.0190</th>
<th>0.0146</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.0187)</td>
<td>(0.0173)</td>
<td>(0.0170)</td>
</tr>
</tbody>
</table>

Student-School Fixed Effects: Y  Y  Y

Observations: 1,064,565  1,064,565  1,064,565

Notes: The cells in this table report estimates from separate regressions. Each regression is estimated at the student level, using all available fourth and fifth grade TPS student-year observations from 1997 to 2011. The estimates in each cell correspond to the effect of post-opening charter school competition within a given distance on student-level math test scores in TPSs. In panel (a), the dependent variable is student-level standardized math scores, and each regression includes the following control variables: year and grade fixed effects; grade-specific cubic polynomials in prior math and reading scores; indicators for student ethnicity, gender, limited English proficiency, and disability status; and classroom, school-grade-year, and school-year averages of prior test scores and all demographic characteristics. Panel (b) includes student-school fixed effects and drops all prior score variables and fixed student-level characteristics from the control vector. Panel (c) maintains the same control vector as panel (b) but changes the dependent variable from level (standardized) math scores to a gain math scores (the difference between the contemporaneous and prior standardized scores). Standard errors clustered at the school level appear in parentheses. *** denotes significance at the 1% level; ** denotes significance at the 5% level; * denotes significance at the 10% level.
Table 1.6: Effects of Competition on Teacher Math Value-Added (VA)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a): Effects of Working in a TPS Facing Competition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working in Comp.</td>
<td>0.0216***</td>
<td>0.0187***</td>
<td>0.0225***</td>
<td>0.0190***</td>
<td>0.0215***</td>
</tr>
<tr>
<td></td>
<td>(0.0045)</td>
<td>(0.0057)</td>
<td>(0.0045)</td>
<td>(0.0066)</td>
<td>(0.0058)</td>
</tr>
<tr>
<td>(b): Effects of Being Observed After Working in Competitive TPS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed Post Comp.</td>
<td>0.0328***</td>
<td>0.0231***</td>
<td>0.0306***</td>
<td>0.0220***</td>
<td>0.0412***</td>
</tr>
<tr>
<td></td>
<td>(0.0055)</td>
<td>(0.0065)</td>
<td>(0.0055)</td>
<td>(0.0074)</td>
<td>(0.0079)</td>
</tr>
<tr>
<td>Observations</td>
<td>88,828</td>
<td>88,828</td>
<td>88,828</td>
<td>62,351</td>
<td>26,477</td>
</tr>
<tr>
<td>Teacher Fixed Effects</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Teacher-School Fixed Effects</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Post-School Switch Controls</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Restricted: Non-Switchers</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Restricted: Switchers</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

Notes: The dependent variable in each column is teacher-year math value-added. Each cell represents an estimate from a separate regression and all regressions are estimated on the full sample of TPS teacher-year observations with an estimated math value-added from 1997 to 2011. Panel (a) presents estimates of the effect of working in a TPS that faces charter school competition within 5 miles. Panel (b) presents estimates of the effect of being observed after working in a TPS facing competition from a charter school within 5 miles. All regressions include teacher fixed-effects, except those in column (2), which include teacher-school fixed effects. Each regression also contains the following control variables: grade and year fixed effects, controls for teacher experience, classroom, school-grade-year, and school-year averages of students’ prior-year math and reading scores, and percentages of minority students, disabled students, limited English proficient students, free-lunch-eligible students (only at the school-year level), and male students. Standard errors clustered at the teacher-level appear in parentheses. *** denotes significance at the 1% level; ** denotes significance at the 5% level; * denotes significance at the 10% level.
Notes: Panels (a) and (b) present estimates of the effects of charter school competition on TPS math and reading value-added, respectively. I report the effects of charter school competition within radii differentiated by half-mile increments. Estimates are derived from separate regressions, identical to those in Tables 1.4. I show the associated 95 percent confidence interval with each point estimate. Additional control variables in each regression include year and school fixed effects and school-year averages of student prior math and reading scores and school-year percentages of minority students, disabled students, limited English proficient students, male students, and free-lunch eligible students.

Figure 1.1: Competitive Effects by Distance from Charter School
Notes: These figures are constructed by estimating school-level event-study equations with all TPS-year observations from 1997 to 2011. The panels report estimates and associated 90 percent confidence intervals for the $\rho_k$ coefficients from variants of equation (1.6). The left out year is the year prior to competition, $k = -1$. Panels (a) and (c) use the full sample of TPSs and TPS math and reading productivity as the dependent variable, respectively. Panels (b) and (d) use the sample of TPSs that ever face charter competition and TPS math and reading productivity as the dependent variable, respectively. Additional control variables in each regression include year and school fixed effects and school-year averages of student prior math and reading scores and school-year percentages of minority students, disabled students, limited English proficient students, male students, and free-lunch eligible students. Standard errors used to construct confidence intervals are clustered at the school level.

Figure 1.2: TPS Math and Reading Value-Added Before and After Charter School Opening
Chapter 1. School Competition, Accountability, and Teacher Productivity

37

Figure 1.3: Effects of Competition on Enrollment, Staff, and Class-size

Notes: These figures are constructed by estimating school-level event-study equations with all TPS-year observations from 1997 to 2011. The panels report estimates and associated 90 percent confidence intervals for the $\rho_k$ coefficients from variants of equation (1.6). The left out year is the year prior to competition, $k = -1$. The dependent variables in panels (a), (b), (c), and (d) are the fraction of third to fifth grade students transferring to a charter school, student enrollment, the number of full-time equivalent teachers, and average class size at a TPS, respectively. Additional control variables in each regression include year and school fixed effects. Standard errors used to construct confidence intervals are clustered at the school level.
Chapter 1. School Competition, Accountability, and Teacher Productivity

Notes: These figures are constructed by estimating school-level event-study equations with all TPS-year observations from 1997 to 2011. The panels report estimates and associated 90 percent confidence intervals for the $\rho_k$ coefficients from variants of equation (1.6). The left out year is the year prior to competition, $k = -1$. The dependent variables in panels (a), (b), (c), and (d) are average prior math scores, average prior reading scores, the fraction of free-lunch eligible students, and the fraction of students with college-educated parents, respectively. Parental education is only available from 1997 to 2006. Additional control variables in each regression include year and school fixed effects. Standard errors used to construct confidence intervals are clustered at the school level. This figure continues with panels (e) through (h) on the following page.

Figure 1.4: Effects of Competition on Student Body Characteristics
Chapter 1. School Competition, Accountability, and Teacher Productivity

39

- .04
- .02
0
.02
.04
.06
Fraction of Students

-4
-3
-2
-1
0
1
2
3
Years Since Charter Opening

(e) Fraction of Disabled Students

- .04
- .02
0
.02
.04
.06
Fraction of Students

-4
-3
-2
-1
0
1
2
3
Years Since Charter Opening

(f) Fraction of Male Students

- .04
- .02
0
.02
.04
.06
Fraction of Students

-4
-3
-2
-1
0
1
2
3
Years Since Charter Opening

(g) Fraction of Limited English Proficient Students

- .04
- .02
0
.02
.04
.06
Fraction of Students

-4
-3
-2
-1
0
1
2
3
Years Since Charter Opening

(h) Fraction of Minority Students

Notes: These figures are constructed by estimating school-level event-study equations with all TPS-year observations from 1997 to 2011. The panels report estimates and associated 90 percent confidence intervals for the $\rho_k$ coefficients from variants of equation (1.6). The left out year is the year prior to competition, $k = -1$. The dependent variables in panels (e), (f), (g), and (h) are the fraction of disabled students, the fraction of male students, the fraction of limited English proficient students, and the fraction of minority students, respectively. Additional control variables in each regression include year and school fixed effects. Standard errors used to construct confidence intervals are clustered at the school level.

Figure 1.4: Effects of Competition on Student Body Characteristics
Continued
Chapter 1. School Competition, Accountability, and Teacher Productivity

Notes: These figures are constructed by estimating school-level event-study equations with all TPS-year observations from 1997 to 2011. Panel (a) shows the evolution of the percentage of teachers who are new hires at a TPS in each year before and after a charter school opens within 5 miles of the TPS. Panel (b) shows the evolution of the percentage of teachers leaving a TPS in each year before and after a charter school opens within 5 miles of the TPS. The figures are constructed by estimating equation (1.6) with the fraction of new hires and the fraction of teachers leaving as the dependent variables and plotting the estimates of the $\rho_k$ coefficients along with the associated 90 percent confidence intervals. The left out year is the year prior to competition, $k = -1$. Additional control variables in each regression include year and school fixed effects and school-year averages of student prior math and reading scores and school-year percentages of minority students, disabled students, limited English proficient students, male students, and free-lunch eligible students. Standard errors used to construct confidence intervals are clustered at the school level.

Figure 1.5: Teacher Turnover Before and After TPSs Face Competition
Notes: These figures are constructed by estimating an event-study equation similar to equation (1.6) at the teacher-level. I first regress teacher-year math value-added on indicator variables for time since the teacher worked in a TPS facing charter school competition. The left out year is the year before the teacher works in such a TPS. I then plot the estimated coefficients on these indicator variables and the associated 90 percent confidence intervals. Panels (a), (b), and (c) each report the results of separate regressions. All regressions include teacher fixed-effects and use TPS teacher-year observations from 1997 to 2011. The regression in panel (a) uses the full sample of TPS teacher-year observations and contains the following additional control variables: grade and year fixed effects, controls for teacher experience, classroom, school-grade-year, and school-year averages of students’ prior-year math and reading scores, and percentages of minority students, disabled students, limited English proficient students, free-lunch-eligible students (only at the school-year level), and male students. The regression in panel (b) uses the same sample, but adds a full set of ‘time-since-switching-schools’ indicator variables as additional controls. Panel (c) removes these additional controls but restricts the sample to teachers who never switch schools. Standard errors used to construct confidence intervals are clustered at the teacher-level.

Figure 1.6: Teacher Math Value-Added Before and After Working in Competition
Chapter 2

Competition and Accountability with (Un)Motivated Teachers
2.1 Introduction

The empirical analysis in the preceding chapter shows that (i) TPSs in North Carolina become more productive at raising student math scores when they face competition from charter schools, and (ii) the TPS-level productivity improvement is driven by a given group of teachers improving performance rather than TPSs hiring new, more productive teachers.

These findings are relevant to the literature that examines how school competition affects teacher labor markets (Hoxby, 2002; Hanushek and Rivkin, 2003; Jackson, 2012). Yet they offer only a partial explanation as to how schools respond to a competitive threat, especially in settings where they face pre-existing accountability programs like the ABCs. The ABCs provide financial bonuses to teachers while increases in competition often do not have direct effects on teachers’ incentives,\(^1\) raising the question how competitive forces cause teachers to exert more costly teaching effort. Most theories of school competition do not address the issue directly, instead positing models in which schools are single decision-making bodies that face additional incentives to raise effort when competition increases (Hoxby, 2003b). In practice, however, rather than being a unitary decision-maker, schools consist of several different teachers whose incentives may differ, and a principal who is tasked with managing them effectively.

Teachers in a given school are likely to be heterogeneous in several ways, including their underlying ability and motivation.\(^2\) Increases in competition then complicate the management problem for school principals, as they are suddenly faced with having to adjust their previous management strategies under the ABCs in order to make a group of heterogenous workers more productive. We currently know little about how the productivity effects of competition differ across teachers – a key omission in the literature, as it is important for understanding how schools coordinate their teachers to improve overall school-level productivity and, correspondingly, the areas in which school-based accountability programs currently provide relatively weak incentives.

This chapter aims to fill this gap by exploring heterogenous effects of competitive pressure

---

\(^1\)It is rare that salaries or potential accountability performance bonuses are tied to total student enrollment.  
\(^2\)There is substantial within-school variation in teacher quality. Chetty et al. (2014a) find that 85 percent of the variation in teacher quality occurs within (not across) schools. In my sample, depending on the measure used, between 83 to 88 percent of the variation in teacher quality occurs within schools.
looking across individual teachers. As a foundation for thinking about how competition affects school principals’ management strategies, I present a model in which teachers are, to varying degrees, ‘motivated agents,’ following the general framework of Besley and Ghatak (2005). Besley and Ghatak (2005) define motivated agents as workers who derive intrinsic benefit from contributing to a successful outcome and use a principal-agent model to show that worker motivation and incentive pay are substitutes for eliciting employee effort. I model teachers as agents who differ in their degree of motivation and show that charter school competition leads to a performance improvement among the least motivated teachers by creating incentives for school principals to monitor these teachers more intensively.

The general model consists of a principal-agent framework in which a school principal manages several teachers. Teachers differ in their degree of motivation, which can be conceptualized as the intrinsic benefit they derive when their school reaches a successful ABCs outcome. A school principal in North Carolina has little control over teacher pay, but in the model she is able to incentivize teachers to exert additional teaching effort by increasing the intensity with which each teacher is monitored. Teachers face a non-pecuniary cost when they are caught shirking on the job, and choose an optimal level of teaching effort by trading off the benefit of contributing to a better school-level ABCs outcome against the costs of effort and the cost of shirking. Principals choose optimal monitoring outlays by weighing the effects of more intense monitoring on teachers’ effort choices against the extra cost.

In the model, motivated teachers never shirk on the job, while unmotivated teachers need extra incentives to exert effort in the form of principal monitoring. When a TPS does not face charter school competition, the teachers and school principal care about how their decisions influence the school’s ABCs outcome; they care relatively little about how these decisions influence school enrollment because students do not have an alternative neighborhood school to

\[3\] Teachers may also derive intrinsic benefit from seeing students perform well. It is difficult to write down a clear objective function for such preferences, however, while the clear rules of school-based accountability systems allow one to write down a transparent objective for schools and teachers.

\[4\] Teachers in North Carolina are paid according to a salary schedule, which is based on experience, education, and National Board Certification status. School principals are able to compensate teachers above the guaranteed amount by creating additional positions for teachers to fill, such as mentor, training, administrative, and coaching positions (Jackson, 2012). I show below that there are no differential salary adjustments across teachers after competition starts.
attend. Thus, the principal chooses to monitor the unmotivated teachers until the additional cost of doing so is exactly offset by the effect of increased teacher effort on the potential school-level ABCs outcome. If the benefit of passing the ABCs to the principal is relatively low, there is an equilibrium in which motivated teachers exert full effort, unmotivated teachers shirk, and the principal does not expend maximal possible outlays monitoring unmotivated teachers.

When a charter school locates near the TPS, the TPS’s students have an alternative schooling option and the school principal must consider how her actions and those of her teachers influence school enrollment and, consequently, total school funding. With her motivated teachers already working at full capacity, if the principal deems that total teaching effort must increase to prevent students from leaving for the charter school, she must expend additional monitoring outlays toward her unmotivated teachers, prompting them to increase teaching effort. Charter school competition changes the benefit of expending monitoring resources for the principal, thus making her more willing to engage with her unmotivated staff, which causes those teachers to exert additional effort. The model yields the intuitive prediction that, if school-level productivity improves, it is driven entirely by performance improvements among the unmotivated teaching staff.

Guided by the model, I classify teachers as motivated and unmotivated according to their level of pre-competition productivity, based on the rich North Carolina data. Then I examine heterogeneity in productivity gains across teachers. Specifically, I use the Empirical Bayes (EB) estimator of teacher value-added (Kane and Staiger, 2008) to estimate pre-competition value-added for each teacher, grouping teachers into quartiles of the value-added distribution, and then estimate differential competitive effects across the quartiles.

Since value-added is measured with noise, mean reversion could confound the analysis of differential effects across teachers. For example, ‘bad’ teachers might improve after competition and ‘good’ teachers might become worse because the former experienced negative performance shocks prior to competition while the latter experienced positive performance shocks. I use two different methods to guard against mean reversion in my analysis. First, using a pooled regression of all teacher-year observations, I restrict attention to teachers who never work in a competitive TPS to estimate the degree of mean reversion directly, thereby isolating the effect
of competition. Second, I rank teachers into quartiles using their value-added from the first year they are observed and then drop the first two observations for each teacher – in other words, the observation used to make the ranking and the one immediately after, during which time mean reversion is likely to be most severe.

Both methods yield very similar results and show that the least-productive teachers prior to competition – those in the first quartile of the value-added distribution – experience the largest productivity improvement as a result of being in a competitive environment. Specifically, these teachers improve by 0.3 standard deviations – enough to move them from the twenty-fifth to the thirty-fifth percentile of the performance distribution. In contrast, teachers in the top quartile experience no change in performance. These results are consistent with competition altering the incentives faced under the ABCs by creating stronger incentives for school principals to expend more resources toward managing their unmotivated teaching staff.

2.1.1 Relation to Prior Literature

This chapter contributes to the literature examining the role of pro-social preferences in employer-employee relationships and to the literatures in the economics of education concerned with estimating teacher productivity and the effects of school competition.

The idea that pro-social preferences or non-pecuniary aspects of motivation are important in employer-employee relationships is well established. The reviews by Gibbons (1998) and Prendergast (1999) document the drawbacks of using standard monetary incentives for eliciting employee effort, chief among them being that such incentives are contingent only on variables that are observable to management and thus vulnerable to gaming by employees. Akerlof and Kranton (2005) argue that many such problems can be mitigated if organizations are able to place workers into jobs with which they identify or help to create of such identities. Employees who identify with the mission of their organization, and therefore derive intrinsic benefit from successful outcomes, are more likely to exert effort without requiring financial bonuses to do so.

The substitutability of intrinsic motivation and monetary incentives for eliciting effort is discussed in several theoretical contexts, including analyses of differences in incentive structures across government and private firms (Francois, 2000), the optimal design of contracts to attract
highly-motivated workers (Delfgaauw and Dur, 2007 and 2008), and the greater flexibility in contract design available to firms with intrinsically-motivated workers (Murdock, 2002).\(^5\)

The theoretical framework in this chapter is most related to Besley and Ghatak (2005), who use the notion of motivated agents to show that optimal incentive pay is decreasing in worker motivation and that, as a result, there are gains to be had from allowing workers to sort into organizations with their most preferred missions. Besley and Ghatak (2005) also argue, quite convincingly, that competition among schools is likely to raise overall productivity because it allows teachers to sort into their most preferred schools, where they exert more effort because of their intrinsic motivation to do so. It is often the case, however, that organizations are not made up of only one principal and one (representative) agent, as is assumed in many previous models. Organizations are also unlikely to be able to recruit and select workers perfectly, resulting in a staff comprised of both motivated and unmotivated agents. The conceptual framework in this chapter therefore differs from previous work by studying how a principal (the school principal) manages several agents (teachers) of differing levels of motivation. The model shows that, in these cases, increases in industry competition can provide incentives for managers to devote more resources toward working with their unmotivated employees.

The empirical literature exploring the importance of pro-social preferences often aims to test whether the alignment of worker preferences and organizational mission can boost productivity without the need for incentive pay. Carpenter and Gong (2016) use a field experiment to show that incentives have minimal effects on workers whose preferences are well-matched with the relevant organizational mission but have large positive effects on the productivity of workers whose preferences are mismatched. Berg et al. (2015) also use a field experiment to show a similar result, in which financial incentives are able to ameliorate the negative effects of having individuals work with others of different social backgrounds. DellaVigna et al. (2016) conduct a model-based field experiment to estimate structural preference parameters, and use the estimates to show that piece-rate incentive pay is critical when social preferences are lacking but that the presence of social preferences can sometimes entirely eliminate the need for incentives.

\(^5\)There is also a literature that suggests monetary incentives can ‘crowd-out’ workers’ intrinsic motivation to perform a task, leading to negative long-run effects. See Benabou and Tirole (2003) for a discussion.
Overall, these experiments suggest that motivation and incentives are indeed substitutes for eliciting worker effort.

In contrast to field experiments, which sort workers into preference-matched and preference-mismatched settings, the empirical analysis in this chapter examines an existing workplace and shows that the least productive workers prior to an increase in industry competition drive the entire average productivity improvement. I argue that this follows from differences in underlying motivation: motivated teachers are the most productive prior to competition because they do not shirk on the job while unmotivated teachers need extra incentives to exert more teaching effort. Thus, the theory and empirical analyses in this chapter work together to identify the employees who are likely to be motivated and unmotivated across several existing workplaces, rather than exogenously sorting workers into motivated and unmotivated roles.\(^6\)

The results in this chapter also contribute to both the literature concerned with estimating the effects of competition in education and the literature that explores options for improving average teacher quality. Studies of school competition typically find only modest aggregate effects of increased competition on student outcomes (Hoxby, 2000; Hoxby, 2003a; Sass, 2006; Booker et al., 2008; Bayer and McMillan, 2010; Card et al., 2010; Neilson, 2013; Figlio and Hart, 2014; Jinnai, 2014). I show that modest aggregate effects can mask substantial productivity effects throughout the distribution of teachers. Competitive incentives are an especially important determinant of productivity for teachers at the bottom of the performance distribution – precisely the teachers whom value-added policies target for evaluation or even dismissal (Hanushek, 2011; Kane and Staiger, 2014; Chetty et al., 2014b).

The remainder of this chapter is organized as follows: The following section sets out a motivated-agents framework in which charter school competition creates incentives for a school principal to combat the differences in underlying motivation of her teaching staff in an attempt to elicit more effort from her relatively unmotivated teachers. Section 2.3 analyzes heterogeneous effects of charter school competition across teachers, showing evidence consistent with

\(^6\)It is possible that differences in underlying performance ability may explain teachers’ initial, pre-competition rankings better than differences in underlying motivation. However, given the marked within-teacher performance improvement at the bottom of the teacher productivity distribution, it is clear that these teachers were previously not working at a level that reflects their full capabilities.
Chapter 2. Competition and Accountability with (Un)Motivated Teachers

the model indicating that the least productive teachers prior to competition show the most improvement after competition. Section 2.4 provides concluding remarks.

2.2 A Model of Teachers as Motivated Agents

2.2.1 Model Setup

The Production Technology and TPS ABCs Payoffs

Suppose that a TPS has $J$ teachers, each of whom have a teaching ability denoted by $a_j$. Also assume that the test score of student $i$, who is assigned to teacher $j$ in year $t$, depends on the student’s prior score (capturing his or her stock of knowledge), teacher ability and effort, and measurement error: $y_{ijt} = \gamma y_{ijt-1} + a_j + e_{jt} - \epsilon_{ijt}$. In this case, the average test score of the students in teacher $j$’s class is given by the sum of the students’ average prior scores, average teacher ability and effort, and average measurement error, $\bar{y}_{jt} = \gamma \bar{y}_{jt-1} + a_j + e_{jt} - \bar{\epsilon}_{jt}$. Aggregating across all teachers, the school-level average score can be written as $\sum_j \omega_{jt}(\gamma \bar{y}_{jt-1} + a_j + e_{jt} - \bar{\epsilon}_{jt})$, where $\omega_{jt} = \frac{n_{jt}}{N_t}$ is the weight assigned to teacher $j$’s class or the ratio of the number of students in teacher $j$’s class to the total number of students at the school.

Letting $\alpha$ denote the parameter governing required student growth, the school’s ABCs score target is $\alpha \sum_j \omega_{jt}\bar{y}_{jt-1}$. The school passes the ABCs when the average score is greater than the required improvement: $\sum_j \omega_{jt}((\gamma - \alpha)\bar{y}_{jt-1} + a_j + e_{jt} - \bar{\epsilon}_{jt}) \geq 0$. Defining $F(\cdot)$ as the cumulative density function of $\sum_j \omega_{jt}\bar{\epsilon}_{jt}$, the probability of the school passing the ABCs may be written as

$$F\left(\sum_j \omega_{jt}(\gamma - \alpha)\bar{y}_{jt-1} + a_j + e_{jt}\right).$$

When the school attains its ABCs target, each teacher and the school principal get paid a financial bonus $b$. I assume the principal values passing the ABCs at a monetary equivalent value of $B \geq b$, as she also takes the school’s reputation into account.
Teachers as Motivated Agents

Besley and Ghatak (2005) set out a principal-agent model in which a more ‘motivated’ agent derives a greater intrinsic payoff, and thus a greater overall payoff, when the project on which they are working results in a successful outcome. Following Besley and Ghatak (2005), I assume that teachers are motivated agents, and that teacher motivation governs the intrinsic payoff a teacher derives from her school successfully passing the ABCs. I specifically assume that there are two types of teacher: those who are ‘motivated’ and those who are ‘unmotivated.’ When her school passes the ABCs, a motivated teacher derives an intrinsic payoff $\theta_h$ while an unmotivated teacher derives an intrinsic payoff $\theta_l$, where $\theta_h > \theta_l$.

The Teacher’s Effort Problem

Teachers are risk-neutral and trade off the benefits and costs of their teaching effort. When the school successfully passes the ABCs, motivated teachers receive a payoff given by $b + \theta_h$ and unmotivated teachers receive $b + \theta_l$. Each teacher affects the probability that the school passes the ABCs by choosing a level of teaching effort, $e_{jt}$, which represents the proportion time she devotes to effective teaching, implying that $0 \leq e_{jt} \leq 1$. For a given level of effort, teacher $j$ incurs a cost of effort given by $c_j(e_{jt}) = \eta e_{jt}^2$.

Following a similar setup as the monitoring model in Dickens et al. (1989), I assume the school principal chooses a monitoring intensity for each teacher given by $m_{jt}$, which results in the principal detecting a teacher not devoting full time to effective teaching with probability $p(m_{jt})$. The detection technology $p(\cdot)$ is increasing and concave, and there is always a positive probability that the principal detects shirking so that $p(0) > 0$. The principal faces constraints over monitoring resources, being able to devote a maximum of $\bar{m}$ in total outlays on monitoring across all teachers, at which point detecting shirking is not guaranteed, $p(\bar{m}) < 1$. When the school principal observes that a teacher is not devoting time to effective teaching, she sanctions the teacher by imposing a cost given by $k$. The sanction cost can be interpreted, for example, as the principal demanding the teacher start completing and submitting weekly lesson plans for review.

The total probability of the principal observing a teacher shirking on the job depends on
both the actions of the principal and the teacher and is given by \( p(m_{jt})(1 - e_{jt}) \). Each teacher takes principal monitoring intensity as given and chooses effort \( e_{jt} \) optimally to maximize the following payoff:

\[
(b + \theta_j)F\left( \sum_j \omega_{jt}((\gamma - \alpha)\bar{y}_{jt-1} + a_j + e_{jt}) \right) - \frac{\eta e_{jt}^2}{2} - kp(m_{jt})(1 - e_{jt})
\]

s.t. \( 0 \leq e_{jt} \leq 1 \).

**The Principal’s Monitoring Problem**

The school principal values passing the ABCs but also considers student enrollment, which determines the overall funding received by the school. Let \( V \) denote the per-pupil funding received by the school for each additional student enrolled. Since school principals are able to observe teacher behavior and performance over time, I assume the principal is fully informed about which teachers are motivated and which are unmotivated. The principal then takes into account teacher behavior and chooses a set of optimal monitoring outlays for her teachers \( \{m_{jt}\}_{j=1}^{J} \) to maximize the following payoff:

\[
BF\left( \sum_j \omega_{jt}((\gamma - \alpha)\bar{y}_{jt-1} + a_j + e_{jt}^*) \right) + N_tV - \sum_j \frac{m_{jt}^2}{2}
\]

s.t. \( \sum_{j=1}^{J} m_{jt} \leq \bar{m} \)

\( m_{jt} \geq 0 \ \forall \ j \)

\( e_{jt}^* \) solves teacher \( j \)'s problem in (2.1) \( \forall j \).

When the principal increases monitoring intensity, she affects the probability of her school passing the ABCs through the effects that additional monitoring has on teacher effort, \( e_{jt}^* \). Additional monitoring is associated with costs given by \( \sum_j \frac{m_{jt}^2}{2} \). When there is no charter school competitor, the school is guaranteed the entire market of students, implying that \( N_t \) does not depend on the actions taken by either the principal or the teachers.\(^7\)

---

\(^7\)In a slight departure from standard principal-agent models, I ignore teachers’ participation constraints. Teaching is usually a heavily unionized or regulated profession, suggesting teachers are likely to be far away from


2.2.2 Effort and Monitoring Decisions without Charter School Competition

To obtain closed-form solutions to the model and to focus on the distinction between the behavior of motivated and unmotivated teachers, I make the following two assumptions.

**Assumption 2.1** The ABCs is a linear piece-rate scheme, instead of a non-linear threshold-based scheme.

Under Assumption 2.1, the cumulative density function $F(\cdot)$ drops out of the objective function of each teacher and the principal, leaving only the current argument of the function being multiplied by the benefit of succeeding on the ABCs. The marginal benefit of effort exertion for each teacher thus depends only on her motivation and the level of principal monitoring; it does not depend on the actions of the other teachers at the school. This greatly simplifies the analysis, eliminating coordination among teachers in their effort choices.

**Assumption 2.2** $\omega_{\min}(b+\theta_h) > \eta > b + \theta_l + k$, where $\omega_{\min}$ is the smallest fraction of students for which a given teacher can be responsible.

Assumption 2.2 is a statement about the degree to which $\theta_h$-type teachers are more motivated than $\theta_l$-type teachers. The marginal benefit from better ABCs performance is always higher than the slope of the marginal cost function for motivated teachers, even in the absence of any principal monitoring. In contrast, unmotivated teachers derive a much lower intrinsic payoff from better ABCs outcomes, leading to a marginal benefit of effort that is always lower than the slope of the marginal cost curve, even with perfect principal monitoring. With these assumptions, the solution to the teacher’s problem is given by Lemma 2.1.

**Lemma 2.1** Under Assumptions 2.1 and 2.2, the solution to the teacher’s problem in equation (2.1) is given by:

$$e_{jt}^* = \begin{cases} \frac{\omega_{jt}(b+\theta_{jt})+p(m_{jt})k}{\eta} < 1 & \text{if } \theta_{jt} = \theta_l \\ 1 & \text{if } \theta_{jt} = \theta_h \end{cases}$$

[2.3]

This assumption is similar to Assumption 1 in Besley and Ghatak (2005), which restricts profits to be less than 1 in order to ensure an interior solution for the agent’s effort level.
The unmotivated teachers choose to spend a smaller fraction of their time dedicated to effective teaching than the motivated teachers, who work at full capacity. Intuitively, this follows from the differences in intrinsic motivation across the two types of teacher: motivated teachers work harder because they derive a higher intrinsic benefit when their school realizes a better outcome under the ABCs. The effort of unmotivated teachers is increasing in the ABCs payoff, \( b \), and decreasing with the cost of effort, \( \eta \). Since they face the threat of sanctions if they are caught shirking on the job, unmotivated teachers also increase effort as the probability of being caught, \( p(m_{jt}) \), increases.

The school principal is aware of how teachers behave and takes this information into account when choosing monitoring levels. Prior to presenting the solution to the principal’s problem in Lemma 2.2, I make the additional simplifying assumption that class sizes are equal within a school.

**Assumption 2.3** Class sizes are equal across all teachers in a given school: \( \omega_j = \omega \forall j \).

Assumption 2.3 helps simplify the principal’s problem by assigning equal weight to the contribution of each teacher’s effort to the school-level ABCs score. With equal class sizes, differences in principal monitoring across teachers are driven only by differences in teachers’ levels of motivation, rather than the behavior of some teachers being relatively more important to the school because they are responsible for more students.\(^9\)

**Lemma 2.2** Let \( J_l \) be the number of type-\( \theta_l \) teachers at the school. Under Assumptions 2.1, 2.2 and 2.3, the solution to the principal’s problem in equation (2.2) is given by:

\[
m^*_{jt} = \begin{cases} 
\frac{m_j}{J_l} & \text{if } \theta_j = \theta_l \& \frac{kp'\left(m_j^*\right)\omega B}{\eta} \geq \frac{m_j}{J_l} \\
\frac{kp'\left(m_j^*\right)\omega B}{\eta} & \text{if } \theta_j = \theta_l \& \frac{kp'\left(m_j^*\right)\omega B}{\eta} < \frac{m_j}{J_l} \\
0 & \text{if } \theta_j = \theta_h
\end{cases}
\]  

\(^9\)Allowing the principal to manipulate class sizes or student-teacher matching greatly complicates the problem and is beyond the scope of this analysis.
The principal optimally chooses not to monitor the motivated teachers, as these teachers work at capacity without additional sanction threats. In contrast, there are gains to be had from monitoring the unmotivated teachers, as these teachers derive a relatively low intrinsic payoff from success under the ABCs and are thus responsive to the threat of being sanctioned if they are caught shirking. Principal monitoring of unmotivated teachers is increasing in the value the principal places on the ABCs, $B$, while it is decreasing in the cost of teacher effort, $\eta$, because teachers are less responsive to the threat of sanctions when they face higher costs of effort. The ratio of $\frac{B}{\eta}$ is critical in determining whether the principal expends maximum monitoring resources: the higher the benefit of passing the ABCs is relative to the cost of teacher effort, the more likely it is that the principal chooses to expend full monitoring resources, which are divided equally across all $\theta_l$-type teachers, $\frac{\bar{m}}{J_l}$.

2.2.3 Effort and Monitoring Decisions with Charter School Competition

When a charter school begins operating nearby, the TPS is no longer guaranteed to capture the entire market of students, $N_t$. Instead, the TPS captures only some fraction of students, which depends on the quality of education offered by the TPS, itself a function of teacher effort. Denote the market share of students captured by the TPS as $Z(q_t)$, where school quality is defined as average teaching quality, $q_t = \sum_j \omega(a_j + e_{jt})$. Making this adjustment to her objective function, the principal again chooses optimal monitoring outlays for each teacher $\{m_{jt}\}_{j=1}^J$ to maximize the following payoff:

$$B \sum_j \omega((\gamma - \alpha)\bar{g}_j + a_j + e_{jt}^*) + Z(q_t)N_tV - \sum_j \frac{m_{jt}^2}{2}$$

subject to

$$\sum_{j=1}^J m_{jt} \leq \bar{m}$$

$$m_{jt} \geq 0 \quad \forall \ j$$

$e_{jt}^*$ solves teacher $j$’s problem in (2.1) $\forall \ j$. 

Proof: See Appendix. ■
The optimal monitoring levels are now given by Lemma 2.3.

**Lemma 2.3** If students value school quality, such that improvements in school quality have non-negative enrollment effects \( \frac{\partial Z}{\partial q_t} \geq 0 \), the solution to the principal’s problem in equation (2.5) is given by:

\[
m^{*}_{jt} = \begin{cases} 
\frac{m}{J_t} & \text{if } \theta_j = \theta_l & \text{if } \theta_j = \theta_l & \text{if } \theta_j = \theta_h \\
\frac{kp'(m^*)}{\eta} (B + N_t V \frac{\partial Z}{\partial q_t}) & \text{if } \theta_j = \theta_l & \text{if } \theta_j = \theta_h \\
0 & \text{if } \theta_j = \theta_h 
\end{cases}
\] (2.6)

**Proof** See Appendix. □

Lemmas 2.1 and 2.3 show that charter school competition potentially affects teachers only indirectly through the effect that it has on principal monitoring decisions. Competition increases the marginal benefit to the principal of monitoring the unmotivated teachers because these teachers increase effort as a result, which positively affects both the ABCs outcome and the school’s total enrollment and funding. The marginal benefit to monitoring motivated teachers does not change, however, as these teachers work at full capacity with or without the threat of sanctions and cannot increase effort above their chosen pre-competition level. Thus, competitive forces only create incentives for the principal to increase monitoring of the unmotivated teachers, and it is these teachers who potentially increase effort when charter school competition starts. Motivated teachers do not change their effort decisions, continuing to work at full capacity.\(^{10}\)

Proposition 2.1 characterizes the full impact of charter school competition on principal monitoring and, correspondingly, teacher and school productivity.

**Proposition 2.1** Let \( m^{*c}_{jt} \) and \( m^{*nc}_{jt} \) denote optimal principal monitoring of teacher \( j \) with

\(^{10}\)The main result does not rely on the assumptions about the exogenous parameters in the utility functions of motivated and unmotivated teachers. A similar result is achieved by allowing motivated teachers also to have interior solutions for effort but introducing curvature to the cost function, making the marginal cost of effort non-linear. In this case, the principal expends non-zero monitoring toward all teachers but monitors motivated teachers less and increases monitoring of unmotivated teachers relatively more when competition starts. The principal’s return to increasing monitoring is higher for unmotivated teachers, as motivated teachers are on a steeper part of the marginal cost of effort curve and therefore respond less to the same change in principal monitoring. See Carpenter and Gong (2016) for a similar discussion.
and without charter school competition, respectively, and let \( e_{jt}^{\ast c} \) and \( e_{jt}^{\ast nc} \) denote teacher \( j \)'s corresponding effort decisions.

**Motivated Teachers.** Charter school competition has no effect on principal monitoring of motivated teachers and has no effect on the effort exerted by motivated teachers.

**Unmotivated Teachers.** If the school principal places relatively high value on ABCs outcomes, such that \( B \geq \frac{m}{J} \frac{\eta}{\omega} \), then charter school competition does not affect principal monitoring of unmotivated teachers and it does not affect teacher effort at the school. If the school principal places relatively low value on ABCs outcomes, such that \( B < \frac{m}{J} \frac{\eta}{\omega} \), then there is an increase from pre- to post-competition principal monitoring of unmotivated teachers given by

\[
m^{\ast c}_{jt} - m^{\ast nc}_{jt} = \begin{cases} \\
\frac{m}{J} \frac{k\eta}{\omega} N_t V \frac{\partial Z}{\partial q_t} \left( B + N_t V \frac{\partial Z}{\partial q_t} \right) \geq \frac{m}{J} \\
\frac{k\eta}{\omega} N_t V \frac{\partial Z}{\partial q_t} \left( B + N_t V \frac{\partial Z}{\partial q_t} \right) < \frac{m}{J}.
\end{cases}
\]

(2.7)

The corresponding increase in unmotivated teacher effort is given by

\[
e^{\ast c}_{jt} - e^{\ast nc}_{jt} = k \left( p(m^{\ast c}_{jt}) - p(m^{\ast nc}_{jt}) \right) > 0.
\]

(2.8)

**Proof** See Appendix.

Proposition 2.1 makes clear that the degrees to which the principal increases monitoring and unmotivated teachers raise effort depend on (i) whether the principal’s total monitoring constraint is binding without charter school competition and (ii) the sensitivity of the school’s enrollment to the quality it offers, governed by the slope of the enrollment curve with respect to school quality \( \frac{\partial Z}{\partial q_t} \).

When the school principal places relatively high value on the ABCs, the principal’s total monitoring constraint is binding prior to competition, implying that the principal cannot increase monitoring any further in response to a charter school competitor. The principal continues to devote her full potential resources toward monitoring \( \theta_t \)-type teachers after competition starts – expending \( m_{jt} = \frac{m}{J} \) units of monitoring toward each of them – and the effort of \( \theta_t \)-type teachers does not increase, as they face no additional incentives to reduce shirking.

In this case, the ABCs creates strong incentives for principals to engage with their unmotivated
teaching staff, leaving little room for competition to strengthen incentives.

When the principal places only moderate value on the ABCs, the monitoring constraint does not bind prior to competition, allowing competitive forces to change principal monitoring and, consequently, teacher effort decisions. Specifically, charter school competition causes monitoring to increase to an amount governed by the sensitivity of the school’s enrollment to changes in school quality \( \frac{\partial Z}{\partial q_t} \). Enrollment is likely to be most sensitive when students place high value on educational quality and the nearby charter school competitors offer a comparable or better level of quality. In these cases, competition causes the principal to expend maximum monitoring resources, leading the level of monitoring devoted to each \( \theta_l \)-type teacher to increase from \( m_{jt}^{nc*} = \frac{kp'(m_{jt}^{nc*})}{\eta} B \) to \( m_{jt}^{c*} = \bar{m}_l \). When the school’s enrollment is not sensitive enough to push the principal maximum monitoring, she instead chooses to only increase monitoring of each \( \theta_l \)-type teacher to \( m_{jt}^{c*} = \frac{kp'(m_{jt}^{c*}) - \omega B}{\eta} (B + N_i V \frac{\partial Z}{\partial q_t}) < \bar{m}_l \).

Although the precise nature of the school’s responses to competition depends on the principal’s valuation of the ABCs and the sensitivity of student enrollment to changes in school quality, the model makes clear that only unmotivated teachers are expected to increase effort in response to competitive forces while the effort of motivated teachers does not change. Teacher ability and effort are both unobserved in the data, but teacher value-added estimates represent a composite measure of both quantities. In the next section, I use these estimates along with variation in effort incentives across teachers to explore heterogeneous performance effects of competition and to empirically test the prediction that only unmotivated teachers increase effort after competition.

## 2.3 Heterogeneity in Competitive Effects Across Teachers

### 2.3.1 Identification Strategy

I test for heterogeneity in the effects of competition across teachers by first categorizing teachers according to their pre-competition value-added in order to identify low- and high-productivity teachers. To do so, I treat the same teacher as a different teacher before she works in a TPS facing competition and after she works in a TPS facing competition, estimating the following
regression across all fourth and fifth grade students from 1997 to 2011:

\[ y_{ijst} = x'_{ijst} \gamma + 1(\text{before } j \text{ in comp}) \times \mu_{j}^{pre} + (1 - 1(\text{before } j \text{ in comp})) \times \mu_{j}^{post} + \epsilon_{ijt}, \quad (2.9) \]

where the dependent variable is the standardized math test score of student \( i \) who is assigned to teacher \( j \) in school \( s \) at time \( t \). The key variable of interest is the pre-competition value-added of each teacher, \( \mu_{j}^{pre} \). As before, teacher value-added is unobserved and must be inferred from average residual test score gains across all students assigned to a given teacher over time. I estimate both pre- and post-competition teacher value-added using the Empirical Bayes (EB) estimator for teacher value-added (see, for example, Kane and Staiger, 2008, and Chetty et al., 2011).11

Having estimated each teacher’s pre-competition productivity, I categorize teachers into quartiles of the \( \mu_{j}^{pre} \) distribution and investigate how the effects of competition differ across teachers in various quartiles. I assume that measured pre-competition teacher quality is increasing in teacher motivation, on average, so that teachers in the lowest quartile of the distribution are the least motivated while teachers in the top quartile are the most motivated. This assumption is valid when teacher ability and motivation are either non-negatively correlated or exhibit a weak negative correlation.

To see this, use the notation of the model above to write teacher-year quality for any teacher \( j \) as \( q_{jt} = a_{j} + \epsilon_{jt} + \overline{\epsilon}_{jt} \), where \( \overline{\epsilon}_{jt} \) is average classroom noise in test scores. For ease of exposition, assume that, by pooling all available years for each teacher, the EB estimator reduces average test score noise to zero. Pre-competition quality of motivated teachers is then \( q_{j} = a_{j} + 1 \) and pre-competition quality of unmotivated teachers is \( q_{j'} = a_{j'} + \overline{\epsilon}_{j'} \), where \( \overline{\epsilon}_{j'} < 1 \) is the average effort exerted by teacher \( j' \) across all sample years. For higher estimates of teacher quality to reflect relatively more motivated teachers, on average, it must be the case that \( \mathbb{E}(a_{j} | \theta_{j} = \theta_{h}) - \mathbb{E}(a_{j} | \theta_{j} = \theta_{l}) > \mathbb{E}({\overline{\epsilon}}_{j} | \theta_{j} = \theta_{l}) - 1 \), which is satisfied when teacher ability and motivation are non-negatively correlated or exhibit a weak negative correlation. I assume

11The control variables include grade and year fixed effects, indicators for teacher experience, grade-specific cubic functions of the student’s prior math and reading scores; indicators for student grade repetition, ethnicity, gender, disability status, and limited English proficiency status; and classroom, school-grade-year, and school-year averages of prior test scores and these demographic variables.
that this condition is satisfied in the data, as it is unlikely that teacher ability and intrinsic motivation exhibit a strong negative correlation.

While the EB estimator is designed to reduce the effects of noise, it cannot eliminate them entirely. Ranking teachers based on a noisy measure of performance makes it difficult to separate the true effects of competition and the effects of mean reversion across the quartiles. If low-productivity teachers perform poorly in the pre-competition period in part because they were unlucky and experienced negative performance shocks, they may experience significant performance improvements in the post-competition period simply as a result of mean reversion in performance. In contrast, mean reversion could lead to a decline in post-competition productivity among teachers who are highly ranked in the pre-competition period.

To make the problem clear, suppose one runs the following regression, where \( p \) denotes the quartiles of the \( \mu^{pre} \) distribution:

\[
\hat{q}_{jt} = \gamma_j + \lambda_t + \sum_{p=1}^{4} 1(p_j = p) PO_{jt} \times \beta_p + \epsilon_{jt}.
\]

The difference-in-differences estimator for the effect of competition in any given quartile \( p \) may be written as

\[
\hat{\beta}_p = \left[ E(\hat{q}|PO_{jt} = 1, p) - E(\hat{q}|PO_{jt} = 0, p) \right] - \left[ E(\hat{q}|PO_{jt} = 0, p) - E(\hat{q}|PO_{jt} = 0, p) \right]
\]

\[
= \beta_p + \left[ E(\epsilon|PO_{jt} = 1, p) - E(\epsilon|PO_{jt} = 0, p) \right] - \left[ E(\epsilon|PO_{jt} = 0) - E(\epsilon|PO_{jt} = 0) \right].
\]

\( \hat{\beta}_p \) captures the true effect of competition on teachers in quartile \( p \) and the difference in the change in noise between teachers in quartile \( p \) and all other teachers who did not enter competition and thus form the control group. Since all teachers are used in the control group, the average change in the control group’s noise is likely to be near zero. This is unlikely to be the case for teachers ranked as either very low- or high-performing in the pre-competition period. Mean reversion is likely to lead to large positive changes in noise for teachers in the first quartile and large negative changes for teachers in the fourth quartile of the \( \mu^{pre}_j \) distribution, resulting
in upward and downward biased estimates of competition, respectively.

I address mean reversion in two ways. My preferred method is to use the sample of teachers who never work in a TPS facing competition to directly estimate the degree of mean reversion in the data and to then isolate the effect of competition. I first randomly assign placebo start of competition dates to teachers who never work in competition, and then use all teachers (both those who face real and placebo competition) to form the distribution of $\mu_{pre}$. Teachers who do eventually work in competition have real pre- and post-competition effects, while placebo teachers have two effects measured in randomly determined periods of time. The ranking of teachers into quartiles then categorizes teachers who actually work in competition based on pre-competition effects and placebo teachers based on their first effects.

Using these rankings and a pooled regression of all teachers, I estimate post-competition effects while simultaneously estimating post-placebo-competition effects in each quartile. Since placebo teachers never face competition, performance fluctuations among them should be entirely driven by mean reversion. Performance of very good placebo teachers should revert back down in the ‘post’ period and performance of very bad placebo teachers should revert back up. These dynamics represent pure mean reversion, suggesting that if any effect is left over for teachers who actually work in competition, it should be attributable to the effects of competition net of mean reversion.

My estimation strategy is to modify the main teacher-level estimating equation from Chapter 1 to the following regression:

$$\hat{q}_{jst} = \sum_{p=1}^{4} 1(p_j = p)\text{Placebo}_{jt} \beta_{p}^{\text{placebo}} + \sum_{p=1}^{4} 1(p_j = p)\text{PO}_{jt} \beta_{p}^{\text{real}} + X_{jst}'\gamma + \lambda_t + \psi_j + \epsilon_{jst}, \quad (2.11)$$

in which the binary variable $\text{Placebo}_{jt}$ is defined as

$$\text{Placebo}_{jt} = \begin{cases} 
\text{PO}_{jt} & \text{if teacher } j \text{ actually faced competition (the ‘real’ post-comp indicator)} \\
1 & \text{if teacher } j \text{ never faced competition and is in the placebo post period} \\
0 & \text{if teacher } j \text{ never faced competition and is in the placebo pre period.}
\end{cases}$$

Since the $\text{Placebo}_{jt}$ indicator nests the true post-competition indicator $\text{PO}_{jt}$, the estimates of
\( \beta_{\text{real}} \) across the different quartiles reflect whatever effect is left over after accounting for mean reversion – that is, the effect of competition, net of mechanical fluctuations in performance.

As a robustness check to the results from estimating equation (2.11), I also use a ‘rank-and-drop’ approach for estimating the heterogenous effects of competition across teachers. Mean reversion is likely to be most severe during the change in teacher performance from the observation used to make the ranking to the one immediately after. I account for this by ranking teachers according to their first teacher-year fixed effect and then estimating equation (2.10) after discarding the first two observations for each teacher – that is, the observation used to make the ranking and the one immediately following it. While this method presents an alternative way to account for mean reversion, it comes with the drawback of only using one year of data to rank teachers, which is likely to result in a less accurate ranking of pre-competition productivity.

### 2.3.2 Results

Table 2.1 presents estimates of differential competitive effects across teachers while Figure 2.1 provides complementary visual evidence. Panel (a) of Table 2.1 shows that, when one does not control for mean reversion through placebo competition, the effects of competition are estimated to be large and positive for teachers in the first quartile and large and negative for teachers in the fourth quartile. In terms of magnitude, the estimated effect among the worst pre-competition teachers is 76 percent of standard deviation in teacher productivity.

Correcting these estimates by including placebo-competition controls in panel (b) shows that most of the effect at the extremes of the pre-competition value-added distribution are driven by mean reversion. The estimate of \( \beta_{\text{placebo}} \) for teachers in the fourth quartile implies that all of the estimated negative effect of competition is due to mean reversion. In contrast, there remain strong positive effects of competition among the least productive teachers prior to competition, as they are estimated to be 0.07 student-level standard deviations more productive at raising student math scores after having worked in a TPS facing competition. This effect amounts to 0.3 standard deviations in teacher productivity, enough to move a teacher from the twenty-fifth to the thirty-fifth percentile of the performance distribution.
Figure 2.1 presents event-study analyses across all quartiles of the pre-competition value-added distribution. There is no evidence of pre-existing trends in productivity driving the heterogenous effects, as the performance of all teachers only changes once they begin working in a TPS facing competition. Panels (a) and (b), and (g) and (h), clearly show, however, how one can easily overstate the effects of competition at the extremes of the teacher performance distribution when the effects of mean reversion are not removed from the effects of competition. In Table 2.2, I present the results obtained from the rank-and-drop approach discussed above. Although teacher rankings using this method are less accurate than those obtained with the EB estimator, the analysis yields virtually identical results.

In sum, these results provide clear evidence that the least productive teachers prior to competition experience the largest performance improvements. These heterogenous competitive effects across teachers are consistent with charter school competition changing how principals manage their unmotivated teaching staff. In the following subsections, I test whether other observable mechanisms drive the results and rule out schools differentially adjusting salaries or class sizes in favor of low-productivity teachers.

Ruling out Differential Salary Adjustments

Teachers in North Carolina are paid according to a salary schedule, which is based on experience, education, and National Board Certification status. School principals are able to compensate teachers above the guaranteed amount, however, by creating additional positions for teachers to fill, such as mentor, training, administrative, and coaching positions (Jackson, 2012). Schools may then motivate low-productivity teachers to work harder after competition starts by differentially increasing their salaries relative to the high-productivity teachers.

I estimate an equation similar to equation (2.11), in which I regress annual salary (in 2014 dollars) at the teacher-year level on the controls in equation (2.11) and the determinants of teacher salary: experience, education, and National Board Certification status. Figure 2.2 presents the results and shows no evidence of low-productivity teachers receiving differential pay increases (relative to high-productivity teachers) once they begin working in a TPS facing competition.
Ruling out Differential Class Size Adjustments

While the analysis in Chapter 1 shows no evidence of changes in average class size at TPSs once a charter school opens nearby, it is possible that the average masks heterogeneity in class size changes across different types of teacher within a TPS. I investigate whether this is the case by estimating equation (2.11) with class size as the dependent variable.

Figure 2.3 presents the results from these regressions and shows no evidence of schools differentially reducing class sizes for low-productivity teachers. In fact, panel (a) shows a small increase in class size among low-productivity teachers. The estimated effect is approximately 0.4 students or 10 percent of a standard deviation in class size. It is therefore unlikely that differential class size adjustments explain the productivity gains at the bottom of the teacher distribution. Indeed, the results are robust to including or excluding class size as a control variable.

2.4 Conclusion

In this chapter, I presented a model of teachers as motivated agents (to varying degrees) along with supporting empirical evidence to help better understand the average productivity effects of competition for schools and teachers that were documented in Chapter 1.

The model provides insight into how charter school competition causes TPSs to improve, despite already facing pre-existing performance incentives under the ABCs. Starting with the basic assumption that some teachers derive more intrinsic benefit from successful ABCs outcomes than others, the model shows that an increase in charter school competition causes TPS principals to increase the amount they expend in monitoring resources differentially across their teaching staff. Specifically, since motivated teachers exert full effort without the threat of sanctions if caught shirking, school principals respond to competition by increasing monitoring of their unmotivated teachers, causing those teachers to respond by increasing teaching effort.

The model predicts that such improvement occurs when the benefit the school assigns to passing the ABCs is sufficiently low. When school-based accountability incentives are not strong enough, school principals are unwilling to engage in costly monitoring of their unmotivated
teaching staff. Competition threatens to reduce overall resources allocated to TPSs, as per-student funding travels with students as they leave for charter schools. The presence of a charter school competitor thus makes TPS principals more willing to monitor unmotivated teachers, leading to increased school-level productivity. Intuitively, the model makes clear how competitive incentives can complement those stemming from school-based accountability programs, compensating in areas where accountability incentives are relatively weak.

The empirical evidence presented is consistent with the model, showing that average teacher productivity gains are driven by large productivity improvements among teachers at the bottom of the pre-competition productivity distribution – that is, those teachers who are likely relatively unmotivated. A teacher in the first quartile of the pre-competition performance distribution is 0.3 standard deviations more effective after working in a competitive environment, an effect large enough to move her from the twenty-fifth to the thirty-fifth percentile of the performance distribution.

While modest aggregate effects of school competition typically cast doubt on the ability of competition-based policies to achieve meaningful changes in student and teacher outcomes, the results in this chapter show that these aggregate effects can mask substantial productivity effects throughout the distribution of teachers. My results also show how the effects of competition can potentially compensate for weaknesses in school-based accountability programs. In North Carolina, pre-existing incentives allow for within-school disparities in teaching effort and charter school expansion puts more pressure on school principals to extract more effort from their unmotivated teachers.

School-based accountability programs are widespread across public education systems in the United States, with the passage of the No Child Left Behind Act of 2001 guaranteeing that schools in each state face some form of accountability. It is unlikely, then, that any future competition-based policies that are enacted, such as private school vouchers or charter school expansion, will be introduced in environments where educators do not already face pre-existing performance incentives. In settings with different accountability regimes than North Carolina’s, competitive incentives and accountability incentives may interact in ways other than those documented in this chapter. For example, the school-based accountability programs in Boston
and New York – two states in which successful charter schools have received media attention – may interact differently with competitive incentives as charter schools expand, leading to different effects on the distribution of teacher productivity. I intend to investigate the effects of competition-based policies in such alternative settings in future work.
## Table 2.1: Heterogeneous Effects of Competition Across TPS Teachers

<table>
<thead>
<tr>
<th>Quartile of Pre Comp VA:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_p )</td>
<td>0.182***</td>
<td>0.071***</td>
<td>-0.002</td>
<td>-0.085***</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.008)</td>
</tr>
</tbody>
</table>

**Notes:** The dependent variable in each column is teacher-year math value-added. Each panel reports the results of different regressions. The cells in each panel represent estimates from the same regression. All regressions are estimated on the full sample of TPS teacher-year observations with an estimated math value-added from 1997 to 2011. All regressions include teacher fixed effects and contain the following control variables: grade and year fixed effects, controls for teacher experience, classroom, school-grade-year, and school-year averages of students’ prior-year math and reading scores, and percentages of minority students, disabled students, limited English proficient students, free-lunch-eligible students (only at the school-year level), and male students. Panel (a) presents estimates of the effects of competition from equation (2.10) using the full sample of teachers and ranking teachers based on the Empirical Bayes estimates of pre-competition value-added. Panel (b) presents estimates of the effects of competition from equation (2.11) using the full sample of teachers and ranking teachers based on the Empirical Bayes estimates of ‘real’ and ‘placebo’ pre-competition value-added.
Table 2.2: Heterogeneous Effects with Rank-and-Drop Method

<table>
<thead>
<tr>
<th>Quartile of Pre Comp VA:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{real}$</td>
<td>0.068***</td>
<td>0.051***</td>
<td>0.039**</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.019)</td>
<td>(0.017)</td>
<td>(0.015)</td>
</tr>
</tbody>
</table>

Teacher-Year Observations 47,448

Notes: The dependent variable in each column is teacher-year math value-added. The cells in each panel represent estimates from the same regression. The regression is estimated on the full sample of TPS teacher-year observations with an estimated math value-added from 1997 to 2011. It includes teacher fixed effects and the following control variables: grade and year fixed effects, controls for teacher experience, classroom, school-grade-year, and school-year averages of students’ prior-year math and reading scores, and percentages of minority students, disabled students, limited English proficient students, free-lunch-eligible students (only at the school-year level), and male students. The cell entries present estimates of the effects of competition from equation (2.10) when teachers are ranked based on their first value-added measures and the first two observations for each teacher are discarded. Standard errors clustered at the teacher-level appear in parentheses. *** denotes significance at the 1% level; ** denotes significance at the 5% level; * denotes significance at the 10% level.

Figures
Chapter 2. Competition and Accountability with (Un)Motivated Teachers

Figure 2.1: Heterogenous Effects of Competition on Teacher Value-Added

Notes: This figure continues with panels (e) through (h) on the following page. Detailed notes are presented there.
Chapter 2. Competition and Accountability with (Un)Motivated Teachers

Notes: These figures are constructed by estimating teacher-level event-study equations with all TPS teacher-year observations from 1997 to 2011. Using teacher-year math value-added as the dependent variable, I estimate modified versions of equation (2.11), in which I replace the $PO_{jt}$ and $Placebo_{jt}$ indicators with a set of ‘time-since-working-in-competition’ and ‘time-since-placebo-competition’ indicator variables. The left out year for teachers in each quartile is the year before the teacher works in competition or starts placebo competition. Each panel plots the estimated coefficients on the ‘time-since-working-in-competition’ indicators along with the associated 90 percent confidence intervals. Panels (a), (c), (e) and (g) are constructed using the same regression, which simultaneously controls for ‘time-since-placebo-competition’ indicators. Panels (b), (d), (f) and (h) are constructed using a different regression, which does not control for ‘time-since-placebo-competition’ indicators. Both regressions include teacher fixed-effects and contain the following control variables: grade and year fixed effects, controls for teacher experience, classroom, school-grade-year, and school-year averages of students’ prior-year math and reading scores, and percentages of minority students, disabled students, limited English proficient students, free-lunch-eligible students (only at the school-year level), and male students. Standard errors used to construct the confidence intervals are clustered at the teacher-level.

Figure 2.1: Heterogenous Effects of Competition on Teacher Value-Added Continued
Figure 2.2: Heterogenous Effects of Competition on Teacher Salary

Notes: This figure continues with panels (e) through (h) on the following page. Detailed notes are presented there.
Notes: These figures are constructed by estimating teacher-level event-study equations with all TPS teacher-year observations from 1997 to 2011. Using teacher-year salary (in 2014 dollars) as the dependent variable, I estimate modified versions of equation (2.11), in which I replace the $PO_{jt}$ and $Placebo_{jt}$ indicators with a set of ‘time-since-working-in-competition’ and ‘time-since-placebo-competition’ indicator variables. The left out year for teachers in each quartile is the year before the teacher works in competition or starts placebo competition. Each panel plots the estimated coefficients on the ‘time-since-working-in-competition’ indicators along with the associated 90 percent confidence intervals. Panels (a), (c), (e) and (g) are constructed using the same regression, which simultaneously controls for ‘time-since-placebo-competition’ indicators. Panels (b), (d), (f) and (h) are constructed using a different regression, which does not control for ‘time-since-placebo-competition’ indicators. Both regressions include teacher fixed-effects and contain the following control variables: grade and year fixed effects, controls for teacher experience, education, National Board Certification status, and classroom, school-grade-year, and school-year averages of students’ prior-year math and reading scores, and percentages of minority students, disabled students, limited English proficient students, free-lunch-eligible students (only at the school-year level), and male students. Standard errors used to construct the confidence intervals are clustered at the teacher-level.

Figure 2.2: Heterogenous Effects of Competition on Teacher Salary
Continued
Figure 2.3: Heterogenous Effects of Competition on Class Size
Notes: These figures are constructed by estimating teacher-level event-study equations with all TPS teacher-year observations from 1997 to 2011. Using teacher-year class size as the dependent variable, I estimate modified versions of equation (2.11), in which I replace the $PO_{jt}$ and $Placebo_{jt}$ indicators with a set of ‘time-since-working-in-competition’ and ‘time-since-placebo-competition’ indicator variables. The left out year for teachers in each quartile is the year before the teacher works in competition or starts placebo competition. Each panel plots the estimated coefficients on the ‘time-since-working-in-competition’ indicators along with the associated 90 percent confidence intervals. Panels (a), (c), (e) and (g) are constructed using the same regression, which simultaneously controls for ‘time-since-placebo-competition’ indicators. Panels (b), (d), (f) and (h) are constructed using a different regression, which does not control for ‘time-since-placebo-competition’ indicators. Both regressions include teacher fixed-effects and contain the following control variables: grade and year fixed effects, controls for teacher experience, classroom, school-grade-year, and school-year averages of students’ prior-year math and reading scores, and percentages of minority students, disabled students, limited English proficient students, free-lunch-eligible students (only at the school-year level), and male students. Standard errors used to construct the confidence intervals are clustered at the teacher-level.

Figure 2.3: Heterogenous Effects of Competition on Class Size
Continued
2.5 Appendix

Proof of Lemma 2.1

Under Assumption 2.1, the teacher’s problem may be written as

\[(b + \theta_j) \sum_j \omega_{jt} ((\gamma - \alpha)\bar{y}_{jt-1} + a_j + e_{jt}) - \frac{\eta e_{jt}^2}{2} - kp(m_{jt})(1 - e_{jt})\]  \hspace{1cm} (2.12)

s.t. \hspace{0.5cm} -e_{jt} \leq 0 \hspace{1cm} (2.12 \text{ i})

\hspace{2cm} e_{jt} \leq 1. \hspace{1cm} (2.12 \text{ ii})

The first-order condition defining optimal effort is given by

\[(b + \theta_j)\omega_{jt} - \eta e_{jt} + kp(m_{jt}) = -\mu + \lambda,\]

where \(\mu\) and \(\lambda\) are the Lagrange multipliers associated with constraint 2.12 (i) and 2.12 (ii), respectively.

We first show that if \(\theta_j = \theta_h\), then \(e_{jt} = 1\). The proof is by contradiction. Suppose not, such that \(e_{jt} < 1\). Then it must be the case that the Lagrange multiplier \(\lambda = 0\), which implies

\[(b + \theta_h)\omega_{jt} - \eta e_{jt} + kp(m_{jt}) = -\mu.\]

By Assumption 2.2, we have \((b + \theta_h)\omega_{jt} - \eta e_{jt} + kp(m_{jt}) > 0\) for all values of \(1 \geq e_{jt} \geq 0\) and for all values \(0 \leq m_{jt} \leq \bar{m}\). But the Lagrange multiplier \(\mu \geq 0\), resulting a contradiction.

We now show that if \(\theta_j = \theta_l\), then \(e_{jt} = \frac{\omega_{jt}(b+\theta_l)+p(m_{jt})k}{\eta} < 1\). The proof is again by contradiction. First, suppose \(\lambda > 0\), so that \(e_{jt} = 1\) and \(\mu = 0\). The first-order condition implies

\[(b + \theta_l)\omega_{jt} - \eta + kp(m_{jt}) = \lambda.\]

By Assumption 2.2, we have \((b + \theta_l)\omega_{jt} - \eta + kp(m_{jt}) < 0\), which is a contradiction. Now
suppose $\mu > 0$, so that $e_{jt} = 0$ and $\lambda = 0$. The first-order condition implies

$$(b + \theta_l)\omega_{jt} + kp(m_{jt}) = -\mu,$$

which is a contradiction because the left-hand side is greater than zero.

Proof of Lemma 2.2

Under Assumptions 2.1 and 2.3, the principal’s problem may be written as

$$B \sum_j \omega_{jt}((\gamma - \alpha)\overline{y}_{jt-1} + a_j + e^*_jt) + N_t V - \sum_j \frac{m_{jt}^2}{2} \tag{2.13}$$

s.t. $\sum_{j=1}^J m_{jt} \leq \bar{m} \tag{2.13 i}$

$-m_{jt} \leq 0 \quad \forall \ j \tag{2.13 ii}$

$e^*_jt$ solves teacher $j$’s problem in (2.1) $\forall \ j. \tag{2.13 iii}$

The first-order condition for each $m_{jt}$ is given by

$$\omega B \frac{\partial e^*_jt}{\partial m_{jt}} - m_{jt} = \lambda - \mu_j,$$

where $\lambda$ and $\mu_j$ are the Lagrange multipliers associated with constraints 2.13 (i) and 2.13 (ii), respectively. Note that, by Lemma 2.1, we have $\frac{\partial e^*_jt}{\partial m_{jt}} = 0$ if $\theta_j = \theta_h$ and $\frac{\partial e^*_jt}{\partial m_{jt}} = \frac{v'(m_{jt})k}{\eta}$ if $\theta_j = \theta_l$.

We first show that if $\theta_j = \theta_h$, then $m_{jt} = 0$. The first-order condition for the principal’s problem for $\theta_h$-type teachers simplifies to

$$-m_{jt} = \lambda - \mu_j.$$

If $m_{jt} > 0$, then $\mu = 0$, but this contradicts with the first-order condition because $\lambda \geq 0$. 

We now show that if \( \theta_j = \theta_l \), then

\[
m_{jt} = \begin{cases} 
\frac{n_j}{m_j} & \text{if } \theta_j = \theta_l \text{ and } \frac{kp'(\frac{n_j}{m_j})\omega B}{\eta} \geq \frac{n_j}{m_j}, \\
\frac{kp'(m_{it}^*)\omega B}{\eta} & \text{if } \theta_j = \theta_l \text{ and } \frac{kp'(\frac{n_j}{m_j})\omega B}{\eta} < \frac{n_j}{m_j}.
\end{cases}
\]

Begin by showing that the Lagrange multipliers \( \mu_j = 0 \) for all \( \theta_l \)-type teachers. The proof is by contradiction. Suppose this is not that case, and that \( \mu_n > 0 \) for some \( \theta_l \)-type teacher indexed by \( n \). There are two cases to consider: \( \lambda = 0 \) and \( \lambda > 0 \).

First, suppose \( \lambda = 0 \). The first-order condition for the principal’s problem associated with teacher \( n \) is then given by

\[B\omega \frac{p'(0)}{\eta} = -\mu_n,\]

which is a contradiction because the left-hand side is greater than zero.

Second, suppose \( \lambda > 0 \). Constraint 2.13 (i) binds, such that \( \sum_{j=1}^{J} m_{jt} = \bar{m} \). Since \( \bar{m} > 0 \), we have that \( m_{it} > 0 \) for at least one \( \theta_l \)-type teacher indexed by \( i \), implying that \( \mu_i = 0 \). The first-order condition for the principal’s problem associated with teacher \( i \) is then given by

\[B\omega \frac{p'(m_{it})}{\eta} - m_{it} = \lambda,\]

and the first-order condition associated with teacher \( n \) is given by

\[B\omega \frac{p'(0)}{\eta} = \lambda - \mu_n.\]

Subtracting one from the other, the two first-order conditions imply

\[B\omega k \left( \frac{p'(m_{it})}{\eta} - \frac{p'(0)}{\eta} \right) = \mu_n + m_{it}.\]

The concavity of \( p(\cdot) \) and \( m_{it} > 0 \) imply the left-hand side is less than zero, but we have assumed \( \mu_n > 0 \) and \( m_{it} > 0 \), thus resulting in a contradiction.

Having established \( \mu_j = 0 \) for all \( \theta_l \)-type teachers, we must consider the solution to the principal’s problem when \( \lambda > 0 \) (binding resource constraint) and \( \lambda = 0 \) (non-binding resource
First, suppose the resource constraint binds and \( \lambda > 0 \). Consider any two \( \theta_l \)-type teachers, \( i \) and \( n \). The first-order conditions for the principal’s problem for these teachers imply

\[
B \omega k \left( \frac{p'(m_{it})}{\eta} - \frac{p'(m_{nt})}{\eta} \right) = m_{it} - m_{nt}.
\]

The strict concavity of \( p(\cdot) \) ensures that the equality is satisfied if and only if \( m_{it} = m_{nt} \) for all \( i \) and \( n \). Combining this with the binding resource constraint implies \( m_{jt} = \bar{m}_J \) for each \( \theta_l \)-type teacher \( j \). Further, \( \lambda > 0 \) implies \( m_{jt} = \bar{m} \) is the solution to the principal’s problem for each \( \theta_l \)-type teacher \( j \) when \( \lambda = B \omega \frac{p'(\bar{m}_J)}{\eta} - \bar{m} > 0 \).

Second, suppose the resource constraint does not bind and \( \lambda = 0 \). In this case, the first-order condition for the principal’s problem associated with any \( \theta_l \)-type teacher \( j \) is

\[
B \omega \frac{p'(m_{jt})}{\eta} = m_{jt},
\]

which establishes the Lemma.

**Proof of Lemma 2.3**

The first-order condition for the solution to the principal’s problem in equation (2.5) for each teacher \( j \) is given by

\[
\omega B \frac{\partial c_j^*}{\partial m_{jt}} + \frac{\partial c_j^*}{\partial m_{jt}} \left( \frac{\partial Z}{\partial q_t} \frac{\partial q_t}{\partial c_j^*} \right) N_t V - m_{jt} = \lambda - \mu_j
\]

For \( \theta_h \)-type teachers we have

\[
-m_{jt} = \lambda - \mu_j
\]

and for \( \theta_l \)-type teachers we have

\[
\omega \frac{p'(m_{it})}{\eta} \left( B + \frac{\partial Z}{\partial q_t} N_t V \right) - m_{jt} = \lambda - \mu_j.
\]
Since $\frac{\partial Z}{\partial q_t} \geq 0$ and is common to all teachers, the proof proceeds in exactly the same steps as the proof of Lemma 2.2 above. ■

**Proof of Proposition 2.1**

**Motivated Teachers.** A comparison of the results in Lemmas 2.2 and 2.3 shows that principal monitoring of motivated teachers is zero with and without competition. The result in Lemma 2.1 shows that motivated teachers always exert full effort, as their effort decisions do not depend on principal monitoring. Thus, competition causes no change to both principal monitoring of motivated teachers and the chosen effort levels of these teachers.

**Unmotivated Teachers.** When the principal places relatively high value on the ABCs, such that $B \geq \frac{\bar{m}}{Jl} \frac{\eta}{\omega k}$, then Lemmas 2.2 and 2.3 guarantee the total monitoring constraint binds both with and without competition. In this case, the change in principal monitoring of any unmotivated teacher is $m_{jt}^{sc} - m_{jt}^{snc} = \frac{\bar{m}}{Jl} - \frac{\bar{m}}{Jl} = 0$. Since the change in unmotivated teacher effort is given by $e_{jt}^{sc} - e_{jt}^{snc} = \frac{k}{\eta} (p(m_{jt}^{sc}) - p(m_{jt}^{snc}))$, and $m_{jt}^{sc} = m_{jt}^{snc}$, it is clear that unmotivated teachers do not increase effort in this case, and there is no change in school quality.

When the principal places relatively low value on the ABCs, such that $B < \frac{\bar{m}}{Jl} \frac{\eta}{\omega k}$, then Lemma 2.2 guarantees the principal chooses to exert $m_{jt}^{snc} = \frac{kp'(m_{jt}^{snc})}{\eta B} \omega B$ units of monitoring toward each unmotivated teacher prior to competition. When competition starts, Lemma 2.3 shows that the principal increases monitoring to $\frac{kp'(m_{jt}^{snc})}{\eta B} \omega B + \frac{kp'(m_{jt}^{snc})}{\eta B} \omega (B + N_l V \frac{\partial Z}{\partial q_t}) < \frac{m}{Jl}$ and to $\frac{m}{Jl}$ if $\frac{kp'(m_{jt}^{snc})}{\eta B} \omega B + \frac{kp'(m_{jt}^{snc})}{\eta B} \omega (B + N_l V \frac{\partial Z}{\partial q_t}) \geq \frac{m}{Jl}$. Taking the difference between post- and pre-competition monitoring establishes the result governing the change in principal monitoring levels. Since post-competition monitoring levels are higher than pre-competition monitoring levels, and the detection probability is increasing in monitoring resources, the change in effort among unmotivated teachers is positive and given by $e_{jt}^{sc} - e_{jt}^{snc} = \frac{k}{\eta} (p(m_{jt}^{sc}) - p(m_{jt}^{snc})) > 0$. ■
Chapter 3

Education Production and Incentives
3.1 Introduction

Assessing the importance of teachers in the production of student achievement has been a central and long-standing preoccupation in academic research. Given that observable teacher characteristics tend to do a poor job of predicting student outcomes, several influential studies – notably, Rivkin, Hanushek and Kain (2005) – have used fixed effects methods to show convincingly that ‘teacher quality’ matters. This finding has led to the development of a substantial related literature, estimating sophisticated value-added measures that seek to capture the overall performance impact of a given teacher (see, for example, papers by Chetty, Friedman and Rockoff 2014a,b). These types of measure have become the focus of policy interventions that include, controversially, firing low value-added teachers, in turn leading to considerable scrutiny of the way value-added measures are constructed.

Alongside the research that estimates teacher value-added, a second active literature has studied the impact of accountability incentives on student and school performance. Accountability schemes have become increasingly prevalent in the United States over the past two decades, sharing the common goal of raising measured teacher and school outcomes by strengthening performance incentives. They range from the federal No Child Left Behind Act of 2001 (‘NCLB’), a proficiency system that sets achievement targets along with penalties for target non-attainment, to various state-level schemes offering rewards or imposing penalties based on the level or growth of student test scores. A number of persuasive papers show that accountability schemes succeed in improving student achievement in a variety of settings.¹

This chapter brings these two literatures together for the first time by drawing a distinction between teacher effects that are invariant to the incentive environment and those that are responsive to it; for convenience, we label the former ‘teacher ability’ and the latter, ‘teacher effort.’² This enables us to address two key research questions: (i) What are the contemporaneous effects of teacher ability versus teacher effort in terms of student scores? (ii) What are

²These labels are not entirely in line with common parlance. We will take effort to be any incentive-related action that raises scores; and while ability is sometimes viewed as being amenable to change (perhaps through teacher training), our measure of ability will equal the component of teacher value-added that does not change over time, conditioning on teacher experience.
the persistent effects of teacher ability versus teacher effort?

In doing so, we move the value-added literature forward in an important new way by separately identifying these two components of teacher effects and establishing that teacher value-added measures are responsive to exogenous incentive variation: as we show, incentive policies are a viable tool for raising teacher value-added. At the same time, our analysis extends the accountability literature, with its emphasis on policy evaluation, by linking incentives explicitly to inputs that enter the education production function. In particular, our estimates of the persistent effects of teacher ability and effort allow us to quantify the relative long-run merits of incentive-based policies alongside policies that focus on teacher selection.

The empirical analysis at the heart of the chapter builds on a simple model of school decision-making. In the model, educators respond to accountability incentives by making effort decisions, which in turn affect student academic performance. Teacher effort and teacher ability are treated as separate inputs in the technology governing student learning, whose effects on future test scores are also allowed to persist at different rates. The model draws attention to cases in which accountability incentives are likely to be particularly strong, providing the basis for the empirical strategies we develop for identifying teacher ability and effort separately and then estimating the persistence of each.

Our empirical analysis uses incredibly rich student-level data covering all public school students in North Carolina from 1997 to 2005. The data allow us to track students and teachers over time and to match students with their teachers accurately in each year. These features of the data are particularly important, as we are able to calculate value-added (“VA”) performance measures for each teacher across several years – critical to our strategy for identifying teacher ability separately from teacher effort. Being able to track students then allows us to explore the persistent effects of effort on future test scores.

In addressing the first question, we exploit plausibly exogenous performance incentive variation to distinguish teacher ability from teacher effort in the year NCLB was introduced in North Carolina. As is well-appreciated in the literature (see Reback 2008, for instance), proficiency schemes like NCLB make students matter differentially at the margin according to their likely test performance relative to a fixed performance target. Drawing on this insight, our strategy
entails comparing teachers who teach classrooms with higher versus lower fractions of marginal students, both before and after the incentive reform was implemented.

Since we observe many teachers before and after, we are able to estimate teachers’ incentive-invariant ability levels using standard VA methods in the pre-reform period. Teacher-effect estimates of this type have been shown to be unbiased predictors of teachers’ average impact on student test scores and important long-run outcomes (see Kane, McCaffrey, Miller, and Staiger 2013, Chetty et al. 2014a, and Kane and Staiger 2014). Once NCLB is implemented, however, we show that these performance measures do in fact vary with incentives in a systematic way, while holding constant estimated teacher ability from the pre-reform period. This approach allows us to separate out the relative importance of teacher ability and effort contemporaneously: a one standard deviation increase in teacher ability is equivalent to 18 percent of a standard deviation increase in student test scores, while an analogous change in teacher effort accounts for 5 percent of such an increase. Given the changes in incentives under the NCLB system, the associated increase in effort is substantial.

To address the second question, we use our contemporaneous measures of teacher ability and effort to investigate the extent to which each input persists in determining future test scores. This component of the analysis is novel in that the literature providing persuasive evidence of the long-run effects of conventional teacher quality measures (e.g. Chetty et al. 2014b) does not allow a role for incentives. To estimate the persistence of teacher ability, we rely on data from the pre-NCLB period and methods used previously in the literature, in which we regress future test scores on standard VA control variables and our measures of teacher incentive-invariant ability. In line with prior work, we find that approximately 40 percent of the initial effect of teacher ability persists after one year and that 20 percent remains after four years.

Estimating the persistence rate of teacher effort is a more challenging exercise. Educators must make two contemporaneous effort decisions in each period after NCLB’s introduction that are potentially confounding: each affects contemporaneous test scores while also being correlated with the initial effort response in 2003. First, educators continue to face NCLB accountability incentives, choosing the amount of effort to devote to each student based on how that student is likely to perform relative to the fixed NCLB performance target. Since
students who are predicted to be on the margin of passing in one year are also likely to be on that margin the next, prior and contemporaneous NCLB effort decisions tend to exhibit strong within-student correlations over time. In order to accurately estimate the persistent effect of effort from the prior period, therefore, we must also account for the test-score effects of contemporaneous NCLB effort, post-2003.

For the second confounding effort channel, schools in North Carolina also face performance targets under the state’s accountability program – the ABCs of public education – both before and after the introduction of NCLB. These targets are set at the school level and depend on students’ prior scores, implying that any effort responses that affect test scores in one period also affect ABCs targets the next. Seen this way, the initial response to heightened NCLB incentives increases schools’ future ABCs targets on average, making it more difficult to meet the passing threshold and potentially causing schools to respond by changing the effort decisions they otherwise would have made. Since these revised effort decisions also affect student test scores, we must hold them constant in order to separately identify the effects of prior-period effort on current test scores from contemporaneous, ABCs-related effort effects.

The ideal experimental design for overcoming these challenges, and thus for cleanly identifying the persistent effects of the NCLB-related effort shock in 2003, would amount to repealing both ABCs and NCLB accountability provisions in 2004 and beyond. In this case, no future effort decisions would be made, allowing us to estimate how effort levels from 2003 (estimated in the first part of our empirical analysis) persisted to affect future test scores using regressions analogous to those used to estimate the persistence of teacher ability.

Because both accountability schemes continued to operate in practice, we use our model to develop a structural estimation approach to control for the contemporaneous test score effects stemming from the two ongoing programs while estimating the persistence of effort from 2003. The approach also accounts for the fact that contemporaneous NCLB effort is a function of the persistence effect that we are trying to estimate.\(^3\)

---

\(^3\)To see this point, recall that proficiency-count systems cause educators to make effort decisions by considering how close to the proficiency target a student is expected to score in the absence of any additional effort. But students’ predicted scores depend on the degree to which the effort they received in the prior period persists forward, thus rendering contemporaneous effort decisions a function of the persistence rate. The parameter governing the persistence of effort thus has to be estimated structurally, as one regressor (contemporaneous
The model emphasizes the dependence of educators’ effort decisions on students’ predicted scores (as noted), implying that effort should be highest among students who are on the margin of passing. We use rich student-level data to show in a non-parametric way that test score gains over predicted scores in 2003 are indeed highest for students who are predicted to score near the passing threshold. The model and the non-parametric estimation patterns allow us to estimate an effort function that takes the distance between students’ predicted scores and the fixed test-score proficiency target as its argument. This function captures NCLB effort levels in both 2003 and 2004, where 2004 effort depends on the degree to which 2003 effort persists. In this way, we account for the first confounding channel.

To account for ABCs-related effort in 2004, we then use the model to show that NCLB’s disruption to the ABCs targets can be well-accounted for by the average effort devoted to all students within a school in 2003. The result follows directly from the ABCs’ design, as the scheme holds schools responsible for average test score growth across all students. We are able to construct average 2003 effort for each school using our estimated effort function above and thus directly control for ABCs effort incentives throughout the estimation routine.

Structural estimates of the persistence of effort reveal that 13 percent of the initial effort effect persists one period forward, which amounts to approximately 25 percent of the one-year persistence of teacher ability. Not accounting for the test-score effects of contemporaneous effort decisions results in an overestimate of the persistence rate of effort: in this case, 50 percent of the initial effort effect is estimated to persist forward, thus highlighting the need for the structural strategy. The faster decay we find for effort is in line with teachers ‘teaching to the test’ to some degree – a phenomenon that is often discussed but rarely identified empirically, though the fact that a portion of prior effort carries over in scores indicates that effort does have longer-term benefits.

These findings are relevant for policy. The prior literature (notably Hanushek 2011 and Chetty et al. 2014b) has focused on altering the teacher ability distribution as a policy lever effort) is inherently a function of the main parameter of interest.

4 We discuss the construction of predicted scores in much greater detail in Section 3.6. Here, we also note that the gains over predicted scores are zero, on average, throughout the entire student distribution in prior years when NCLB incentives do not operate, strongly suggesting that the observed response in 2003 reflects NCLB effort.
to raise student scores. Specifically, policy proposals have featured the controversial notion of replacing teachers whose value-added falls in the bottom five percent of the measured distribution. Building on our finding that incentives matter when measuring the effects of teachers, changing formal incentives constitutes an alternative way of raising student and school performance. Specifically, we propose an incentive-based reform that is more cost-effective than the leading ability-based reform considered in the literature, costing 12 percent less on a per-teacher basis for the same long-run gain.

The remainder of the chapter is organized as follows: The next section describes institutional background to the setting we consider, including the accountability programs we rely on for identification and the rich data used in our analysis. Section 3.3 provides motivating descriptive evidence. Section 3.4 sets out a simple framework for analyzing the production of student achievement that provides the basis for our empirical approach. Section 3.5 outlines the empirical strategy we employ to decompose ability and effort contemporaneously and also presents the results of this exercise, and Section 3.6 presents our strategy for estimating the persistent effects of effort (versus ability) along with the associated results. Section 3.7 draws out policy implications from the analysis; and Section 3.8 concludes.

3.2 Institutional Background and Data

We conduct our analysis in North Carolina, a state that provides significant variation in performance incentives across teachers and schools as well as rich longitudinal data covering all public schools, and their teachers and students. We discuss each in turn.

3.2.1 Accountability Incentives

On the institutional side, the state offers useful incentive variation arising from two separate accountability regimes. The first of these, North Carolina’s ABCs of Public Education, was implemented in the 1996-97 school year for all schools serving kindergarten through grade eight. Under the ABCs, each grade from three to eight in every school is assigned a grade-specific growth target, which depends on both average previous student performance and a constant
level of expected growth. Based on average school-level gains across all grades in student standardized mathematics and reading scores, the ABCs pay a monetary bonus to all teachers and the principal if a school achieves its overall growth target.\(^5\)

Provisions under NCLB — the second of the accountability regimes operating — were implemented in North Carolina in the 2002-03 school year,\(^6\) following the passage of the federal No Child Left Behind Act in 2001. In contrast to the ABCs’ rewards-based approach, NCLB sets penalties for under-performing schools. The program categorizes students into nine subgroups and requires schools to ensure that the percentage of students in each subgroup who achieve proficiency status on the relevant state test meets the state-mandated target. If a school fails to meet any of its subgroup-specific targets, it faces an array of penalties that become more severe over time in the event of repeated failure.\(^7\)

### 3.2.2 Data and Descriptive Statistics

Alongside these accountability regimes, North Carolina provides incredibly rich longitudinal education data from the entire state, available through the North Carolina Education Research Data Center (NCERDC). These data contain yearly standardized test scores for each student in grades three through eight, and encrypted identifiers for students and the teachers who proctor their tests, as well as unencrypted school identifiers. Thus, students can be tracked longitudinally, and linked to a teacher and school in any given year.

Our sample runs from 1997-2005 and is substantial, covering over 2.5 million student-year observations. In terms of performance measures, it includes end-of-grade (EOG) test score performance data for mathematics and reading for all third to eighth grade public school students in the state, although we focus on students in third to fifth grade, for whom we can accurately construct teacher VA estimates. We also observe a ‘pre-test’ in grade three, which is written at the beginning of the year and is treated as the grade two baseline test for third graders.

Table 3.1 provides summary statistics for the unrestricted sample of students. In the analysis

\(^5\)For more detailed descriptions of the ABCs program, see Vigdor (2009) and Macartney (2016).
\(^6\)We will refer to an academic year by its second calendar year. On that basis, NCLB provisions took effect in 2003.
\(^7\)A more detailed description of NCLB can be found in Ahn and Vigdor (2014).
below, we construct our ability and effort measures for each teacher by using individual student test scores. These are measured on a developmental scale, designed so that each additional point represents the same amount of knowledge gained irrespective of the baseline score and school grade. Both the mathematics and reading scores in the table show a monotonic increase across grades, consistent with knowledge being accumulated in those subjects over time. The test score levels are relevant under NCLB, which requires that a given proportion of each of the nine student subgroups (referred to above) exceeds a target score on standardized tests.

The longitudinal nature of the data set enables us to construct growth score measures for both mathematics and reading, based on within-student gains. Those gains are positive, on average, in both subjects across grades, though the largest gains occur for both subjects in the earlier grades. Student gain scores are, as noted, the focus of the ABCs program, which sets test score growth targets for schools, requiring that students demonstrate sufficient improvement as they progress through their educational careers.

The data set includes information about individual students’ gender, race, disability status, limited English-proficiency classification, free lunch eligibility, and grade progression. In the aggregate, about 40 percent of students are minorities (non-white), 6 percent are learning-disabled, only 3 percent are limited English-proficient, and 44 percent are eligible for free or reduced-price lunch. Around 25 percent of students have college-educated parents, and very small fractions of students repeat a grade. These demographic characteristics serve as control variables in our analysis below.

Given our interest in exploring the separate effects of teacher ability and effort, we need to match students in the EOG files to their teachers in an accurate way in any given year. We construct the sample used in our analysis by following previous studies that use the NCERDC data, restricting attention to students in third through fifth grade (as mentioned), where the teacher recorded as the test proctor tends to be the teacher who taught the students throughout the year. We also follow Clotfelter, Ladd, and Vigdor (2006) and subsequent research by only counting a student-teacher match as valid if the test proctor in the EOG files teaches a self-contained class for the relevant grade in the relevant year and if at least half of the tests administered by that teacher are for students in the correct grade. Special education and honors
classes are excluded from the analysis, but we retain students who repeat or skip grades.

### 3.3 Motivating Descriptive Evidence

In this section, we present suggestive evidence that helps to motivate the more structurally-oriented analysis that follows. First, we provide clear evidence of an incentive response consistent with there being a targeted effort increase predicted by standard theory; second, we provide correlational evidence indicating for the first time in the literature that teacher value-added is responsive to incentives.

Our strategy for uncovering teacher effort draws heavily on the introduction of NCLB in 2003, treating that as an exogenous shock to educators’ performance incentives. It has already been well-established that proficiency-count systems like NCLB provide educators with strong incentives to direct resources to students who are on the margin of passing relative to the fixed target, potentially at the expense of those in the tails of the predicted test-score distribution.\(^8\) Thus, the introduction of NCLB should give rise to an inverted U-shaped increase in test scores, peaking where students are directly at the margin of passing.

That prediction is captured empirically in Figure 3.1, which shows that this was exactly the response that occurred in North Carolina when NCLB was introduced in 2003. The figure indicates that gains over predicted test scores in 2003 were highest for students predicted to score close to the test-score proficiency threshold. Gains over predicted scores are decreasing as one moves further away from the threshold, consistent with the predicted effort responses to proficiency-count programs.

In contrast, also shown in the figure, students experienced essentially no gain over their predicted scores at any point of the predicted score distribution in the pre-NCLB year. The flat profile in the pre-NCLB period lends credence to the notion that the incentive shock was exogenous.

Next, building on the notion that NCLB provided differential incentives to exert greater effort depending on how marginal students were relative to the fixed score target, we provide

---

\(^8\)See, for example, Reback (2008), Ladd and Lauen (2010), and Neal and Schanzenbach (2010).
correlational evidence indicating that teacher-year fixed effects, which are commonly used to measure teacher effectiveness, actually depend on a simple proxy for NCLB incentive strength in 2003.\(^9\) We start with a proxy for incentive strength, defining a student as ‘marginal’ if she is predicted to score within \(+/-4\) developmental scale points of the proficiency cutoff,\(^10\) and calculate the fraction of students in each classroom who are marginal. Then, since teacher-year VA represents average residual student test score gains within a classroom, teacher-year VA should be an increasing function of average student NCLB incentive strength within the relevant classroom.

This is what we find, as Figure 3.2 shows. Teacher-year fixed effects depend positively on the proportion of marginal students within a classroom in 2003. The relationship is significant (at the one percent level) and positive in each grade in 2003, with a one standard deviation increase in the proportion of marginal students within a classroom being associated with a 7 percent, 17 percent, and 11 percent standard deviation increase in teacher-year VA in grades three, four, and five, respectively. These raw-data patterns are suggestive of NCLB causing teachers to exert more effort in a manner directed by the incentives involved. In contrast, in the pre-NCLB period, we would expect there to be no relationship between the proportion of marginal students in a classroom and teacher VA.\(^11\)

Next, we set out a simple theory that clarifies the identification challenges that arise in our setting and provides the basis for our empirical strategy, before discussing our strategy for separately identifying teacher ability and effort in Section 3.5.

### 3.4 Theory

In this section, we present a model that allows us to express the two research questions in terms of accountability incentives and parameters that relate to a production technology for student

---

\(^9\)The estimation of teacher-year fixed effects is discussed in much greater detail in Section 3.5 below.  
\(^10\)The results are robust to several choices of cutoff for defining marginal students.  
\(^11\)This is also what we find, though more than simple correlational plots are required in order to account for a confounding negative correlation between marginal student presence and teacher ability. In Section IV.B, we estimate incentive-invariant teacher ability and account for potentially differential sorting of students to teachers in 2003, along with the effects of teacher experience and random classroom shocks on performance to ensure that the relationship in Figure 3.2 indeed reflects a teacher effort response.
learning, thus motivating our empirical analysis. The model also helps motivate our approach to identifying teacher ability and effort separately and also estimating the persistence of each – issues we address later in the section.

3.4.1 The Environment

There are three main elements to the model: the education production technology; the teacher effort decision, made in light of prevailing accountability incentives; and teacher assignments (taken as given). Several other aspects are suppressed, including student effort, student peer effects, and parental help.

Production Technology

We begin with the technology governing education production. As the exact specification of the technology is unknown and many inputs are unobserved, researchers typically limit analyses to linear production technologies and to inputs that are observed in school administrative data sets. Furthermore, most studies examine the effects of contemporaneous inputs, accounting for the history of past inputs by controlling for prior test scores.

We focus on both the contemporaneous and dynamic effects of two unobserved inputs, namely teacher ability and effort. As a starting point, we abstract from other inputs and consider the following production technology, which makes explicit how each of the two inputs affects student learning, both contemporaneously and over time:

\[ y_{ijgst} = \beta_i h_i + \sum_{0 \leq t' \leq t} \left[ \gamma_{a_{ij(g(i,t'))g(i,t')s(i,t')}} a_{ij(g(i,t'))g(i,t')s(i,t')} + \gamma_{e_{ij(g(i,t'))g(i,t')s(i,t')}} e_{ij(g(i,t'))g(i,t')s(i,t')} \right] + \nu_{ijgst}. \]  

(3.1)

Equation (3.1) describes the test score of student student \( i \) who is assigned to teacher \( j \) in grade \( g \) at school \( s \) and time \( t \). Here, \( h_i \) is the (innate) human capital of student \( i \), \( a_{ij(g(i,t'))g(i,t')s(i,t')} \) is the ability of the teacher to whom student \( i \) was assigned in time \( t' \) when the student was in grade \( g(i,t') \) at school \( s(i,t') \), \( e_{ij(g(i,t'))g(i,t')s(i,t')} \) is the effort choice of that teacher at that time, and \( \nu_{ijgst} \) is an error term.

To simplify the analysis, we equate the contemporaneous effects of teacher ability and
effort, assuming $\gamma^a_{t,t'} = \gamma^e_{t,t'} = 1$. Our empirical goals are to separately identify teacher ability and effort, $a_j(i,t')g(i,t')s(i,t')$ and $e_j(i,t')g(i,t')s(i,t')$, and their effects on test scores, $\{\gamma^a_{t,t'}; \gamma^e_{t,t'}\}_{t'=t}$.

Separate identification of ability and effort requires observing a teacher in at least two different time periods in which she faces different effort incentives – a condition that is satisfied in our setting by the introduction of NCLB into a state with pre-existing incentives. Further, credible estimates of the persistent effects of ability and effort (in answering our second research question) are of key policy interest, as these parameters help determine the relative long-run benefits of policies aimed at altering the distribution of teacher ability and those aimed at altering effort incentives for a given group of teachers.

As is the case in all studies of teacher effects, we must account for non-random sorting of students to teachers in order to estimate teacher effects in a credible way. We do so by following previous work and controlling for prior student test scores. To clarify the link between the technology just outlined and our subsequent empirical analysis, one can subtract the $\gamma$ multiplied by the prior score from both sides of the test score equation:

$$y_{ijgst} - \gamma y_{ijg-1s't-1} = a_j(i,t')g(i,t')s(i,t') + e_j(i,t')g(i,t')s(i,t') + (\beta_t - \gamma \beta_{t-1})h_i + \cdots$$

$$+ \sum_{0 \leq t' \leq t-1} \left[ \left( \gamma^a_{t,t'} - \gamma^a_{t-1,t'} \right) a_j(i,t')g(i,t')s(i,t') + \left( \gamma^e_{t,t'} - \gamma^e_{t-1,t'} \right) e_j(i,t')g(i,t')s(i,t') \right] + (\nu_{tgjst} - \gamma \nu_{ijg-1jst-1}).$$

(3.2)

The resulting expression for test scores, with the relevant (simplifying) relabeling, is

$$y_{ijgst} = \gamma y_{ijg-1s't-1} + a_j + e_{ijgst} + \epsilon_{ijgst}.$$  

(3.3)

Here, measurement error in test scores $\epsilon_{ijgst}$ contains the effects of all prior test score inputs, unless the decay rates of teacher ability, teacher effort, and the effects of innate student human capital are all equal to $\gamma$. The simplified production technology implies that the test score of student $i$, who is assigned to teacher $j$ in grade $g$ and school $s$ in year $t$, depends on his or her prior score (or accumulated stock of knowledge), his or her teacher’s ability, the effort the teacher chooses to devote to the student, and a random shock to performance. The persistence
rate of the prior score is a composite rate, capturing the persistent effects of teacher ability, teacher effort, and baseline student human capital.

**Effort Decision**

The fact that neither teacher ability nor teacher effort is observed poses a challenge for estimation: these components must be estimated from residual variation in student tests score. We will use the fact that we observe the same teachers in different incentive environments, thus allowing us to identify ability and effort separately and also trace out the effects of each over time. Our approach will rely on an exogenous shock to performance incentives for teachers created by the introduction of NCLB in 2003.

Here, we first describe how the incentive environment prior to NCLB, under the state’s ABCs program, results in relatively uniform effort incentives across teachers, leaving little variation with which to identify teacher effort. We then show that NCLB’s shock to pre-existing incentives generates systematic variation in incentives across teachers, proving helpful in identifying teacher effort separately from teacher ability.

**ABCs:** The ABCs program sets growth targets that are grade-and-subject-specific for each school and then aggregates across all grade-subject pairs within a school to form a school-level growth score. Let $G_s$ denote the highest grade served by school $s$. The average growth targets in a subject-grade are a linear function of students’ prior scores. Let $\alpha$ denote the (pre-determined) coefficient that multiplies student prior scores to form these targets. School $s$ passes the ABCs when the sum of the differences between average and target scores in each grade is greater than zero:

$$
\sum_{g=3}^{G_s} \sum_{i: i \in gst} \frac{y_{isgt} - \alpha y_{ijg-1st-1}}{N_{gst}} \geq 0.
$$

(3.4)

Here, the sum is taken over all students who are in grade $g$ in school $s$ at time $t$ and $N_{gst}$ is the number of students in that grade in school $s$ at time $t$. 
Substituting the production technology from equation (3.3) into equation (3.4) yields

\[ \sum_{g=3}^{G_s} \bar{\epsilon}_{gst} \geq \sum_{g=3}^{G_s} \left( (\alpha - \gamma) \bar{y}_{g-1st-1} - \bar{a}_{gst} - \bar{e}_{gst} \right) , \]  

(3.5)

where the upper bars denote school-grade-year-specific averages. We let \( F(\cdot) \) denote the cumulative density function of \( \sum_{g=3}^{G_s} \bar{\epsilon}_{sgt} \) and \( b^{abcs} \) denote the financial bonus associated with passing the ABCs. We further assume that the cost of effort for each student \( i \) is given by the strictly increasing and convex function \( c(e_{ijgst}) \). The school is modelled as a centralized decision-maker, choosing a sequence of effort levels across all students \( \{e_{ijgst}\}_{i=1}^{N_{gst}} \) in each year to maximize the following expected-utility payoff:

\[ b^{abcs} \left( 1 - F\left( \sum_{g=3}^{G_s} \left( (\alpha - \gamma) \bar{y}_{g-1st-1} - \bar{a}_{gst} - \bar{e}_{gst} \right) \right) \right) - \sum_{g=3}^{G_s} \sum_{\{i: i \in gst\}}^{N_{gst}} c(e_{ijgst}) , \]  

(3.6)

To simplify the notation moving forward, we define \( \Gamma_{st} \equiv \sum_{g=3}^{G_s} \left( (\alpha - \gamma) \bar{y}_{g-1st-1} - \bar{a}_{gst} - \bar{e}_{gst} \right) \). The first-order condition for the optimal effort devoted to any student \( i \) is given by

\[ b^{abcs} f(\Gamma_{st}) \frac{1}{N_{gst}} = c'(e_{ijgst}) , \]  

(3.7)

where \( f(\cdot) \) is the probability density function of \( \sum_{g=3}^{G_s} \bar{\epsilon}_{sgt} \). Equation (3.7) shows that the only source of within-school effort variation is differential grade sizes, \( N_{gst} \), implying that the ABCs results in relatively uniform levels of effort across students within a given school. Since student characteristics do not vary much over time within schools, we follow Macartney (2016) by assuming that schools reach steady-state levels of ABCs effort. The ABCs environment thus results in minimal within-school variation in teaching effort; indeed, much of the variation is found across schools and is driven by variation in the steady-state likelihood of ABCs success, denoted by \( f(\Gamma^*_s) \).

Schools with differing values of \( \Gamma^*_s \) are also likely to differ along unobserved student and teacher characteristics that determine student test scores, making it challenging to identify
the test score effects of teacher effort. Even if accountability pressure under the ABCs was randomly assigned across schools, the relatively small within-school incentive variation created by the scheme leaves little variation with which to separately identify teacher ability and effort.

NCLB: We will rely on the introduction of NCLB as an exogenous disruption to incentives that provides useful within-school variation in incentives strength across teachers.

Starting in 2003, schools face incentives under NCLB in addition to the ABCs. NCLB sets test score proficiency targets that are fixed across students in a given grade; and it requires that a predetermined percentage of students within a school achieve proficiency status. We abstract from NCLB’s subgroups here, instead assuming that only the school-level pass rate is relevant for success or failure under NCLB.\(^\text{12}\)

Letting \(y_{T,g,nclb}^T\) denote the test score proficiency target in grade \(g\), using the production technology in equation (3.3), and letting \(H(\cdot)\) represent the cumulative density function of \(\epsilon_{igst}\), the probability that student \(i\) reaches proficiency status under NCLB may be written

\[
1 - H(y_{T,g,nclb}^T - \gamma y_{ig-st}^{g-1} - a_j - \epsilon_{igst}).
\]

Aggregating across all students, the school’s expected pass rate becomes

\[
R_{st} = \sum_{g=3}^{G_s} \sum_{i \in g_{st}} \frac{1 - H(y_{T,g,nclb}^T - \gamma y_{ig-st}^{g-1} - a_j - \epsilon_{igst})}{N_{st}}.
\]

We follow Neal and Schanzenbach (2010), assuming that the school’s NCLB payoff is increasing in the fraction students who are proficient (independently of how that fraction compares with the required school-level proficiency rate) and approximating the total number of students who pass as the expected number who pass. We also let \(\Psi(\cdot)\) denote the strictly increasing

\(^{12}\text{We do not model subgroups directly, which is a reasonable first-order approximation to the incentives we exploit: holding fixed whether or not a given subgroup is held accountable under NLCB, the marginal benefit of effort for educators within that subgroup is always higher for students who are predicted to score near the proficiency threshold, irrespective of whether the subgroup is actually accountable. While students within accountable subgroups receive more effort, schools face incentives to devote effort to all marginal students, as marginal students who are not members of accountable subgroups in the current year may be members of accountable subgroups in future years.}\)
concave reward function that takes the fraction of students who pass as its argument.\footnote{Neal and Schanzenbach (2010) model the expected fail rate and thus have a strictly increasing and convex penalty function.}

Combining the model of NCLB with that of the ABCs above, the school’s objective function in 2003 becomes

$$U_{st} = b^{abcs}(1 - F(\Gamma_s^*)) + \Psi(R_{st}) - \sum_{g=3}^{G_s} \sum_{i \in g_{st}} c(e_{ijgst}), \quad (3.10)$$

and the first-order condition for the effort devoted to any student $i$ is

$$b^{abcs} f(\Gamma_s^*) \frac{1}{N_{gst}} + \Psi'(R_{st}) \frac{h(y_{g,nclb}^T - \gamma y_{ijg'} - 1 + a_j - e_{ijgst})}{N_{gst}} = c'(e_{ijgst}). \quad (3.11)$$

This simple model makes clear how the effort devoted to each student within a school now has both an individual and common component. The individual component comes from NCLB. Under this scheme, a student’s position in the student ability distribution determines how much effort he or she receives as dictated by the function $h(y_{g,nclb}^T - \gamma y_{ijg'} - 1 + a_j - e_{ijgst})$. This is the only source of within-school variation in effort across students, which we rely on heavily in our identification strategy for separating teacher ability and effort.

### 3.4.2 Building Intuition for Separating Ability and Effort

Our first research question involves separately identifying teacher ability and effort. The model suggests a natural way to decompose traditional teacher-year VA measures – representing average residual test scores across all students within a given classroom – into teacher ability and effort components. To see this, further assume that the cost of effort is given by $c(e) = \frac{d}{2}e^2$. In
this case, the VA estimate for teacher $j$ in 2003, $q_{j,2003}$, may be written as

$$q_{j,2003} = \frac{N_{j,2003}}{\sum_{i=1}^{N_{j,2003}}} \frac{y_{ijgst} - \gamma y_{ij'g-1's't-1}}{N_{j,2003}}$$

$$= \frac{N_{j,2003}}{\sum_{i=1}^{N_{j,2003}}} a_j + e_{ijgst} + e_{ijgst}$$

$$= a_j + \frac{\int \text{abc} f(\Gamma_s^*)}{dN_{gst}} + \frac{\Psi'(R_{st})}{dN_{st}} \sum_{i=1}^{N_{j,2003}} h(y_{g,nclb}^{T} - \gamma y_{ij'g-1's't-1} + a_j + e_{s}^{*})$$

$$+ \frac{N_{j,2003}}{\sum_{i=1}^{N_{j,2003}}} \frac{e_{ijgst}}{N_{j,2003}}, \quad (3.12)$$

where the third equality follows from substituting in the first-order condition for effort (equation (3.11)). Teacher VA measures in 2003 thus capture teachers’ incentive-invariant ability levels, $a_j$, baseline ABCs effort, average NCLB-related effort across all students in the teacher’s class, and classroom average noise in test scores.

Assuming the distribution of noise is symmetric around a mean of zero, NCLB effort is highest for students who are predicted to score closest to the passing threshold without additional NCLB effort – that is, those students for whom the distance between the test score proficiency target, $y_{g,nclb}^{T}$, and $\gamma y_{ij'g-1's't-1} + a_j + e_{s}^{*}$ is relatively small in absolute terms, where $e_{s}^{*}$ indicates steady-state ABCs effort. Students who are predicted to score far above the threshold are likely to pass without any additional help while students who are predicted to score far below require a prohibitively costly amount of effort to change their expected proficiency status. Thus, the marginal benefit of extra effort for educators is highest among students who are predicted to be on the margin of passing.

It follows that teacher VA in 2003 is increasing in the fraction of these marginal students in a given a classroom, reflecting the relationship between NCLB incentives and school effort decisions. In Section 3.5, we use this fact to generate exogenous variation in effort incentives across teachers, allowing us to identify teacher ability and effort separately by comparing teacher performance in 2003 to that in prior years when NCLB incentives do not exist.
3.4.3 Building Intuition for Estimating the Persistence of Effort

Once we identify the amount of NCLB effort devoted to each student in 2003, our second research question involves estimating how the effects of effort persist forward in 2004. Since both NCLB and the ABCs continue to operate in 2004, this exercise involves separating the persistent effects of 2003 effort from (i) the effects of contemporaneous NCLB effort and (ii) school-level effort responses to potentially disrupted ABCs targets.

To make the identification challenges clear, we decompose the sources of teacher effort given by the first-order condition in equation (3.11), writing the total effort received by student \( i \) as the sum of ABCs- and NCLB-related effort:

\[
e_{ijgst} = e^{\text{abcs}}_s + e^{\text{nclb}}_{ijgst}. \]

Here, \( e^{\text{abcs}}_s \) represents steady-state ABCs effort, which does not vary within schools, and \( e^{\text{nclb}}_{ijgst} \) represents student-specific NCLB effort, which is determined by how close to the proficiency threshold student \( i \) is predicted to score. In 2003, student test scores are thus given by

\[
y_{ijgst} = \gamma y_{ij'g-1's't-1} + a_j + e^{\text{abcs}}_s + e^{\text{nclb}}_{ijgst} + \epsilon_{ijgst}.
\]

In 2004, we allow prior-period NCLB effort to persist at a potentially different rate than pre-NCLB test score inputs, which persist at rate \( \gamma \). Test scores in 2004 are written as

\[
y_{ijgst} = \gamma(y_{ij'g-1's't-1} - e^{\text{nclb}}_{ij'g-1's't-1}) + a_j + e^{\text{abcs}}_s + e^{\text{nclb}}_{ijgst} + \gamma e^{\text{nclb}}_{ij'g-1's't-1} + e_{ijgst}, \tag{3.13}
\]

where \( \gamma^e \) denotes the persistence rate of NCLB effort from 2003. ABCs effort is denoted as \( e^{\text{abcs}}_s \) to indicate that it is no longer necessarily the steady-state level of effort chosen by school \( s \) prior to NCLB, owing to NCLB’s potential disruption of ABCs targets.

There are two key challenges to empirically identifying the persistence of effort, \( \gamma^e \). First, NCLB-related effort decisions in 2004 are a function of NCLB effort from 2003 and the persistence rate. To see this, note that the student-specific term in the first-order condition for effort (equation 3.11) becomes

\[
h(T_{g,nclb} - \gamma(y_{ij'g-1's't-1} - e^{\text{nclb}}_{ij'g-1's't-1}) - a_j - e^{\text{abcs}}_s - e^{\text{nclb}}_{ijgst} - \gamma e^{\text{nclb}}_{ij'g-1's't-1}) \]

in 2004. The persistence rate of prior-year effort determines student human capital in 2004 and, correspondingly, how close to the test-score proficiency target a student is expected to score. Educators thus make contemporaneous student effort decisions in 2004 based in part on the degree to which they believe prior effort will persist.
Second, the persistence rate of prior effort also potentially affects schools’ ABCs incentives. In particular, write the probability of a school passing the ABCs in 2004 as

\[
1 - F \left( \sum_{g=3}^{G_s} \left( (\alpha - \gamma)\bar{y}_{g-1st-1} + (\gamma - \gamma^e)\bar{\epsilon}_{g-1st-1}^{\text{nclb}} - \bar{a}_{gst} - \bar{\epsilon}_{gst}^{\text{abc}} - \bar{\epsilon}_{gst}^{\text{nclb}} \right) \right)
\]  

(3.14)

If \( \lambda > \gamma^e \), NCLB effort from the prior year decreases the probability of ABCs target attainment in 2004, holding all else equal. This is due to the discrepancy between the actual persistence of effort and the rate assumed by the ABCs target. The target takes the full prior score into account, not discriminating between potentially differential persistence rates of the inputs that contribute to that score. If effort persists at a lower rate than non-effort inputs, however, test scores in 2004 should increase at a lower rate than they would in the pre-NCLB period, when non-effort inputs were responsible for all test score gains. Since the ABCs target holds schools to the same standard as in the pre-reform period, the same level of test score improvement from the year \( t-1 \) makes it more difficult for schools to reach contemporaneous ABCs targets in 2004 relative to pre-NCLB years.

It is also clear that average 2003 NCLB effort is the key determinant of the distortion to 2004 school-level ABCs incentives. There is no distortion only when effort persists at the same rate as all non-effort inputs, such that \( \gamma = \gamma^e \). Otherwise, responses to NCLB in 2003 have direct implications for schools’ ABCs effort choices in 2004, as schools respond to the changed likelihood of meeting ABCs targets. However, it is important to note that the first-order conditions defining optimal ABCs effort continue to be the same across all students within a given school, implying that ABCs incentives continue to vary only across schools. We make use of these observations as a way to control for the effects of ABCs effort while estimating the persistence of NCLB effort.

Estimating the persistence of effort thus requires a strategy that allows us to account for contemporaneous effort incentives from both NCLB and the ABCs. We outline our approach in Section 3.6, showing how a combination of non-parametric estimation along with structural assumptions informed by our model can be used to identify the persistence rate of effort separately from contemporaneous incentives.
3.5 Separating Contemporaneous Effort and Ability

This section addresses our first research question, describing our identification strategy for decomposing a teacher’s contribution to her students’ test scores into ability and effort components followed by our findings.

3.5.1 Identification Strategy

We start by noting that standard teacher VA methods are not well-suited for this decomposition. Those methods construct average residual test score gains across all classrooms taught by a given teacher, relying on several observations for the same teacher over time to minimize the influence of noise. For a given teacher VA estimates reflect that teacher’s incentive-invariant ability and the average effort she devotes across all of her students over many years. If effort incentives exhibit substantial within-teacher variation, VA methods cleanly identify teachers’ incentive-invariant ability levels but the variation in effort decisions over time implies that the effects of effort are averaged out. In contrast, when incentives do not vary much over time for a given teacher, VA estimates reflect a composite of both teacher ability and average effort. In neither case do VA measures distinguish explicitly between the two components of teacher effectiveness.

Our chosen approach involves a three-step procedure. By way of overview, we first compute teacher-year fixed effects for each teacher from 1997 to 2005. (This step is largely in keeping with existing approaches taken in the literature.) Second, we use pre-reform data to identify the sum of incentive-invariant ability and pre-existing baseline (ABCs) effort using the Empirical Bayes estimator of teacher VA (Kane and Staiger, 2008; Chetty et al., 2011). Third, we estimate NCLB-induced teacher effort, using the estimated teacher fixed effects from 2003 along with estimates of the sum of teacher ability and baseline effort, and the fraction of students in a teacher’s classroom who are ‘marginal’ with respect to the NCLB target.\textsuperscript{14} We now describe each step in greater detail.

1. Teacher-Year Fixed Effects: When estimating teacher VA, recent studies typically

\footnote{We define a student as ‘marginal’ if she is predicted to score between four points below the proficiency threshold and four points above it. Alternative definitions can easily be considered.}
control for high-order polynomials in previous math and reading scores along with students’
demographic characteristics. We follow that convention, regressing contemporaneous math
scores on cubics in prior math and reading scores and several other student characteristics, in
addition to the teacher-year fixed effects that are our primary focus.

While standardized (usually at the grade-year level) test scores are the predominant out-
come measure in the literature, we opt to measure test scores on a developmental scale instead.
Standardizing test scores guards against changes in testing regimes over time, but de-meaning
effectively removes the effects of aggregate changes in performance incentives. As our pri-
mary interest is assessing how teacher effort affects student learning, we prefer to preserve all
incentive-driven performance variation over time and therefore measure test scores on the de-
velopmental scale throughout our analysis. Here, we rely on the careful psychometric design of
these scales, which ensures that one can track improvements or declines in learning as students
progress through school.

We define the control vector, $x_{ijgst}$, to include student race, gender, disability status, limited
English-proficiency classification, parental education, and an indicator for grade repetition.
The inclusion of these variables helps mitigate the bias stemming from non-random student-
teacher matching, as they are likely correlated with innate student ability and previous teacher
assignments. A teacher-year fixed effect is calculated for a teacher only if she has greater than
seven but fewer than forty students in her class with valid test scores and demographic variables
in the relevant year.\footnote{We exclude a student from the value-added analysis if any of the following criteria are satisfied: (1) multiple
scores for current or lagged EOG math or readings tests; (2) EOG scores corresponding to two or more teachers
in a given year; (3) EOG scores corresponding to two or more grades in a given year; (4) or EOG scores
corresponding to two or more schools in a given year.}

To estimate teacher-year fixed effects, we use the full sample from 1997 to 2005 and run the
following grade-specific regressions (for third, fourth and fifth grades):\footnote{We estimate the teacher-year fixed effects using Stata’s ‘areg’ command. ‘areg’ solves the co-linearity problem
arising from including a fixed effect for each teacher-year pair by estimating each teacher-year effect relative to
an arbitrarily selected constant (equal to the average teacher-year fixed effect in the case of constant class size).
Since the regressions are grade-specific, teacher-year effects produce a ranking of teachers over time (1998 to
2005) within a given grade. For more details, see McCaffrey et al. (2012).}

$$y_{ijgst} = f(y_{ij}^{g-1}s^{t-1}) + q_{jt} + x'_{ijgst} \beta + \epsilon_{ijgst},$$

(3.15)
obtaining the teacher-year fixed-effects estimates as

\[
\hat{q}_{jt} = \sum_{i=1}^{n(j,t)} \frac{y_{ijgst} - \hat{f}(y_{ijg-1s't-1}) - x_{ijgst}^T \hat{\beta}}{n(j,t)} = a_j + \frac{t_{abc} f(\Gamma_{gst})}{N_{gst}^{2003}} + 1_{nclb} e_j \sum_{i=1}^{N_{gst}^{2003}} \frac{h(y_{g,nclb}^T - \gamma a_j y_{ijg-1s't-1} - a_j - e_{igst})}{N_{j2003}} + \bar{\epsilon}_{jt}
\]

\[
= a_j + \epsilon_j + 1_{nclb} e_j (m_{jt}) + \bar{\epsilon}_{jt}
\]

(3.16)

where the second equality follows from equation (3.12) in the model above and \(1_{nclb}\) is a binary variable indicating the post-NCLB period. Teacher-year fixed effects consist of incentive-invariant ability and, in North Carolina, they also consist of baseline ABCs effort, which is denoted as \(e_j\) throughout the empirical analysis. Recall that ABCs effort does not vary across teachers within a school, implying that VA methods largely capture variation in incentive-invariant ability prior to NCLB. Once NCLB takes effect in 2003, the composition of students in a teacher’s classroom helps to determine effort incentives. In particular, since schools want to devote more effort to students who are predicted to score near the test score proficiency standard, the fraction of these marginal students in the classroom, \(m_{jt}\), is likely to influence teacher performance in 2003 and beyond.

The incentive to devote effort to marginal students – a feature common across all proficiency-count systems – allows us to separately identify NCLB-related effort from classroom-specific shocks to test scores in 2003, \(\bar{\epsilon}_{jt}\). We cannot simply relate the teacher-year fixed effects from 2003 to some function of the fraction of marginal students within a classroom, however, as we must also address sorting patterns of students to teachers. Specifically, the proficiency cutoff is set relatively low in North Carolina, implying that marginal students are typically low-performing students. Since low-performing students tend to have low-ability teachers, there is likely to be a negative correlation between teacher ability and marginal student presence within a classroom. We address this issue by estimating teacher incentive-invariant ability in the pre-NCLB period and controlling for it directly throughout the analysis.

2. **Incentive-Invariant Ability and Baseline Effort:** Consistent with much of the existing literature, we assume incentive-invariant ability is fixed over time, conditional on teacher
experience.\textsuperscript{17} We assume baseline ABCs-related effort is fixed also, as ABCs incentives only vary at the school level and most schools reach steady-state levels of effort prior to NCLB.\textsuperscript{18} For each teacher $j$, we estimate the combination of her fixed ability and baseline effort by employing the Empirical Bayes ("EB") estimator of teacher VA in the pre-NCLB period (Kane and Staiger, 2008; Chetty \textit{et al.} 2011). Specifically, we estimate the following pooled specification across all grades and years from 1997 to 2002,\textsuperscript{19} in which we regress test scores on grade-specific cubic polynomials of prior scores, indicators for student ethnicity, gender, limited-English proficiency, disability status, parental education, grade repetition, grade and year fixed effects, and controls for teacher experience:\textsuperscript{20}

\begin{equation}
y_{ijgst} = g(y_{ijg-1s't}) + x'_{ijgst} \beta + h(expjt) + \psi_{ijgst}, \text{ where} \\
\psi_{ijgst} = \mu_j + \theta_{jt} + \epsilon_{ijgst}, \text{ and } \mu_j = a_j + e_j.
\end{equation}

The EB estimator uses several years of data for each teacher to construct an optimally-weighted average of classroom-level residual test scores in order to separate teacher ability, $\mu_j$, from classroom-specific shocks, $\theta_{jt}$, and student-level noise, $\epsilon_{ijgst}$. While we implement the same procedure as previous studies, we note that in our setting, EB estimates consist of both incentive-invariant ability and baseline ABCs effort, $\hat{\mu}_j = \hat{a}_j + \hat{e}_j$.

3. \textbf{NCLB-Induced Effort Response:} As previously mentioned, we define a student as ‘marginal’ if she is predicted to score within $\pm 4$ developmental scale points of the proficiency cutoff. For each classroom, the relative incentive strength measure, $m_{jt}$, is defined as the fraction of students in that classroom who are marginal. We then identify the component of teacher-year quality that is attributable to NCLB effort incentives by regressing teacher-year fixed effects

\begin{itemize}
\item For estimators that allow teacher ability to ‘drift’ over time, see Goldhaber and Hansen (2013), Chetty \textit{et al.} (2014a), Rothstein (2014), and Bacher-Hicks \textit{et al.} (2014).
\item Assuming that baseline effort is fixed over time in our setting is reasonable, given that the pre-existing value-added incentive scheme is approximately uniform in its effects on teacher effort (see Macartney \textit{et al.} (2015)).
\item Due to the differential timing of the 2001 math developmental scale change in third grade, and the unavailability of second grade scores in 1996, we run a separate regression for third grade from 1998 to 2000.
\item We paramaterize the experience function by including indicators for each level of experience from zero to five years, with the omitted category being teachers with six or more years of experience. We choose this specification to be consistent with Chetty \textit{et al.} (2014a). Wiswall (2013) shows that such a restrictive choice may bias estimates of the dispersion in teacher quality – an issue we intend address in future work.
\end{itemize}
\( \hat{q}_{jt} \) on \( m_{jt} \), while holding constant teacher incentive-invariant ability (and baseline effort) and teacher experience:

\[
\hat{q}_{jt} = \alpha + \rho m_{jt} + \lambda(a_j + \epsilon_j) + w(exp_{jt}).
\] (3.18)

Once accountability pressure due NCLB is introduced (captured by \( m_{jt} \)), teachers exert additional effort, according to the amount of pressure they face. We thus test whether there is no systematic relationship between \( \hat{q}_{jt} \) and \( m_{jt} \) prior to NCLB but a positive relationship in 2003. We identify effort as a predicted value from equation (3.18), \( e(m_{j2003}) \), representing the response to the new incentive scheme and capturing the relationship (conditional on ability and experience) between teacher-year performance and the classroom fraction of marginal students in 2003:

\[
e(m_{j2003}) = \hat{\rho} m_{j2003}.
\] (3.19)

### 3.5.2 Results for Research Question One

#### Identifying Ability and Effort

Table 3.2 reports means and standard deviations of the main teacher-level variables of interest, while Figure 3.3 presents the incentive-invariant teacher ability distribution,\(^{21}\) where incentive-invariant ability is defined as the EB estimate from equation (3.17). The distributions are similar among third, fourth, and fifth grade teachers, with means of -0.07, -0.09 and -0.06 developmental scale points, and standard deviations of 1.68, 1.37 and 1.3 developmental scale points, respectively. The EB estimator shrinks teacher effects toward the mean to minimize the influence of measurement error, thus resulting in observed standard deviations that are downward-biased. The estimated ‘true’ standard deviations reported in Table 3.2 are slightly larger, at 2.16, 1.63 and 1.63 developmental scale points. Averaged across grades, mean teacher ability is -0.074 scale points and its standard deviation is 1.79 scale points, or equivalently 0.18

\(^{21}\) This includes the baseline effort level discussed in the prior section, which is also assumed invariant to NCLB incentives.
student-level standard deviations. (The latter is within the upper-bound of the range found by most previous work that estimates teacher quality.)

The top panels of Figure 3.4, (a) through (c), show the grade-specific partial relationships between $\hat{q}_{jt}$ and $m_{jt}$ from equation (3.18), while panel (a) of Table 3.3 reports the underlying estimates of $\rho$. For each grade, we plot the relationships that prevail in 2003 and a pooled regression of all pre-NCLB years that additionally includes year fixed effects. In 2003, there is a clear increasing relationship between the part of the teacher-year effect unexplained by ability and experience and the proportion of marginal students in the classroom. Relative to the 2003 raw-data patterns found in Figure 3.2, we add plots of the corresponding relationships in pre-NCLB years, noting that our three-part procedure accounts for any correlation between incentive-invariant ability and NCLB incentives. Conditional on teacher ability and experience, a one standard deviation increase in the proportion of marginal students within a classroom causes a 9 percent, 22 percent, and 16 percent standard deviation increase in teacher-year VA in third, fourth, and fifth grade, respectively. As expected, there is virtually no relationship in the pre-NCLB years, once we account for the negative correlation between teacher ability and the classroom fraction of marginal students.\footnote{When controlling for teacher ability in the pre-NCLB period, we apply the jack-knife EB estimator, which uses information from all other years than the one in question (Chetty et al., 2011). This avoids mechanical correlation in measurement error driving any of the results.}

Incentives under NCLB cause modest \textit{average} improvements in student test scores: a one standard deviation increase in the fraction of marginal students in a classroom corresponds to test score improvements of 0.03, 0.06, and 0.04 student-level standard deviations in third, fourth, and fifth grade, respectively. We show further the relevance of these incentives by documenting that they lead to within-teacher performance \textit{improvements}. To that end, we construct the difference between 2003 and 2002 teacher-year fixed effects as

$$\hat{q}_{j2003} - \hat{q}_{j2002} = a_j + e_{j2003} + \bar{e}_{j2003} - (a_j + e_{j2002} + \bar{e}_{j2002}) = e_{j2003} - e_{j2002} + \bar{e}_{j2003} - \bar{e}_{j2002}, \quad (3.20)$$

and estimate the relationship between this difference and the fraction of marginal students that teachers faced in 2003, $m_{j2003}$. To ensure that mean reversion is not responsible for the results,
we also control for a cubic function of 2002 teacher-year VA. Panels (d) through (f) of Figure 3.4 show the partial relationships between performance improvement in 2003 and $m_{2003}$, while panel (b) of Table 3.3 reports the underlying slope coefficients. As expected, within-teacher performance improvements are clearly increasing in the fraction of marginal students within classrooms in 2003. The transition from 1999 to 2000 is used as a placebo control in each grade, revealing a flat relationship and strengthening the claim that the 2003 patterns reflect teachers improving performance as a result of NCLB effort incentives.

Panels (g) through (i) of Figure 3.4 present the full distributions of effort in each grade in 2003, where effort is constructed as the fitted value from equation (3.19). The last three columns in Table 3.2 report the associated means and standard deviations. Mean effort for third, fourth and fifth grades is 0.56, 0.8, and 0.45 developmental scale points, respectively (the average across grades is 0.61 points). Dispersion in effort across grades is 0.27, 0.64 and 0.36 points, and is 0.48 scale points across all grades, which amounts to 0.05 student-level standard deviations. Taken together, the evidence indicates that there is meaningful variation in teachers’ effort in 2003, strongly correlated with the fraction of students in their classes who are marginal.

**Addressing Potentially Differential Sorting of Students to Teachers**

To infer teacher effort, we compare the relationship that prevailed between a teacher’s performance and the fraction of students in her class who are marginal in 2003 with the relationship that prevailed in previous years. Since we see a positive relationship in 2003 and no relationship in prior years, we argue that the 2003 relationships reflect changes in teachers’ effort, drawing on the theoretical predictions concerning responses to proficiency-count systems.

A competing explanation is that students were *differentially* sorted to teachers in 2003 such that high (incentive-invariant) ability teachers received larger fractions of marginal students. While we control for teacher ability in the process of estimating effort responses, if high-ability teachers were better able to respond to the demands of NCLB, we might worry that this non-linear relationship between teacher and student ability is driving the results rather than additional effort being exerted by a *given* teacher.
A natural way to evaluate this rival hypothesis is to test whether the relationship between the fraction of marginal students in a classroom and teacher ability changes in 2003. We conduct this test by regressing the fraction of marginal students in each class on grade and year fixed effects, our measure of combined teacher incentive-invariant ability and baseline effort, and an interaction of that term with a year-2003 indicator:

$$m_{jt} = \alpha_0 + \lambda_g + \lambda_t + \beta_1(\hat{a_j} + \hat{e_j}) + \beta_2(\hat{a_j} + \hat{e_j}) \times 1(t = 2003) + \epsilon_{jt}, \forall t \leq 2003.$$ (3.21)

If principals began differentially sorting students to teachers on the basis of ability in 2003, we would expect to find the coefficient on ability and baseline effort not equal to zero ($\beta_2 \neq 0$).

Table 3.4 shows the results from estimating variants of equation (3.21). Overall, there is a small negative relationship between the fraction of marginal students who are in a teacher’s class and that teacher’s incentive-invariant ability. This reflects the relatively low test score proficiency standard in North Carolina and the sorting of historically low-performing students to low-ability teachers. The estimates in column (1) imply that a one standard deviation better than average teacher has 0.61 percentage points fewer marginal students in her class (which corresponds to a 2.3 percent reduction relative to the mean fraction). The sorting patterns appear to change slightly in 2003, but in a way that high-ability teachers receive smaller fractions of marginal students than in the pre-NCLB period. This change is in the opposite direction to that required to bias our results.

While we are able to isolate the relationship between teacher performance and NCLB incentives conditional on teacher ability, the low test score proficiency target and the sorting patterns of students to teachers result in slightly stronger effort incentives for low-ability teachers. Figure 3.5 shows grade-specific relationships between $e(m_{jt})$ and $a_j + e_j$. For each grade, we plot the relationships that prevail in 2003 and the pooled pre-NCLB control years. In 2003, there is a clear decreasing relationship between NCLB effort and incentive-invariant ability. In the control years, the estimated functions are virtually flat.

The lack of a relationship between placebo NCLB effort and teacher ability is due to NCLB incentives not operating in those years: $e(m_{jt})$ is identified by holding ability constant, implying...
that we automatically account for the correlation between ability and $m_{jt}$ in the estimation routine. Since $m_{jt}$ is not associated with any incentives prior to NCLB, there is no remaining relationship to identify. In contrast, in 2003, $m_{jt}$ reflects NCLB incentives, and the analysis reveals stronger effort responses among the teachers who faced stronger incentives. The slope coefficients are all significant at the one percent level and are $-0.01$, $-0.04$, and $-0.02$, for third, fourth and fifth grade, respectively. Thus, a one standard deviation lower ability teacher in 2003 exerted approximately 0.02, 0.07, and 0.03 developmental scale points worth of additional effort in third, fourth grade, and fifth grade. These differences are very small, however, corresponding to 0.002, 0.007, and 0.003 student-level standard deviations.

### 3.6 Estimating the Persistence of Teacher Ability and Effort

In this section, we address our second research question – whether teacher ability and effort persist at different rates. As in the previous section, we outline our estimation strategy in some detail before presenting the empirical results.

#### 3.6.1 Estimating the Persistence of Ability

We begin by estimating the persistence of ability and baseline effort in a reduced-form way, following the previous literature (for example, Chetty et al., 2014b). Specifically, we regress student test scores in period $t+n$ on the full control vector from the EB regression (equation (3.17)) and the ability of the teacher who taught the students in period $t$:

$$y_{ijgst+n} = f_g(y_{ijg-1s}') + x_{ijgst}' \beta + h(exp_{jt}) + \beta_n(a + \varepsilon)_{jt} + \epsilon_{ijgst}. \quad (3.22)$$

Here, ability is measured with a jack-knife EB estimator, which uses information from all years except the current year to form an estimate of teacher ability. This prevents mechanical correlation between measurement error in test scores and in teacher ability from confounding the results (Chetty et al. 2014a). The coefficient $\beta_n$ represents the degree to which the effect of teacher ability from year $t$ persists forward to influence test scores in year $t+n$.

Figure 3.6 presents the estimated $\beta_n$ coefficients from regressions that include test scores
exclusively from the pre-NCLB period. As a check, teachers do not affect their students test scores in the years before they are matched with these students, as shown by the estimate at \( t - 2 \). As expected, a one unit better-than-average teacher in year \( t \) improves student test scores by 1 developmental scale point, on average. The contemporaneous effect of teacher ability fades away over time, as 41 percent of the initial effect persists to affect test scores in period \( t + 1 \), and only 20 percent remains by period \( t + 4 \).

These results align closely with the prior literature (Jacob, Lefgren, and Sims, 2010; Chetty et al. 2014a). Having estimated the one-year ahead persistence of ability at 0.41, we seek to obtain estimates of the persistence of teacher effort and contrast the two persistence effects. Estimating the persistence of effort is a more challenging exercise, as the strong within-student correlation of NCLB effort over time implies that one must account for contemporaneous effort to avoid overstating the persistent effects of lagged effort. Yet, because effort persistence helps determine student predicted scores, and educators make effort decisions based on anticipated student performance, contemporaneous effort is itself a function of the persistence parameter. We also must take into account potential changes to ABCs effort to avoid confounding school-level ABCs-related improvement with student-level effort persistence.

We now discuss the identification issues in greater detail and develop a strategy for estimating the persistence of effort.

### 3.6.2 Estimating the Persistence of Effort

#### Measuring Student-Level Effort

In the analysis above, we identified teacher effort at the classroom level. This is a natural measure for assessing if teacher VA depends on performance incentives in a systematic way, as teacher-year VA itself represents classroom-level residual test score gains. Classroom effort is not our preferred measure for estimating persistence, however, as more variation in the data can be exploited with a student-level measure. Measuring effort at the classroom level may over- or under-state the effort received by each student, leading us to a framework for measuring effort

---

23 Since we control for once-lagged test scores when estimating teacher ability, the coefficient at \( t - 1 \) is identically zero.
at the student level and investigating the rate at which student-specific effort persists.\footnote{It is well-established that much of the variation in student test scores occurs within rather than across schools \citep{KaneStaiger2002}, implying that much of the variation in students' predicted test scores also occurs within schools. Since NCLB proficiency targets are \textit{fixed} across students, and the distance between predicted scores and the target determines student-level incentive strength, it follows that much of the variation in educators' effort incentives occurs within schools. We therefore prefer to use a student-level measure of effort in order to make use of all of the available variation.}

We measure student-level effort by building on the non-parametric patterns in Figure 3.1, which show that the introduction of NCLB had large effects on student test scores, in a way that is consistent with a strong teacher effort responses to the scheme.\footnote{More details are available in \citet{Macartney2015}.} The incentive strength measure on the horizontal axis is constructed in three steps. In the first step, we predict student performance in a flexible way in pre-NCLB years using several covariates, including lagged test scores.\footnote{Specifically, we regress contemporaneous 2002 scores on cubic in prior 2001 math and reading scores and indicators for parental education, gender, race, free or reduced-price lunch eligibility, and limited English proficiency.} In the second step, we use the saved regression coefficients from the first step to construct a predicted score for each student in 2003, denoted by $\hat{y}_{ijgs2003}$, by combining those coefficients with updated (but pre-determined) student covariates. In the third step, we use the NCLB target, $y_{Tg,nclb}$, to compute incentive strength as the difference between the predicted score and the target, written as $\hat{y}_{ijgs2003} - y_{Tg,nclb}$\footnote{The predicted score is invariant to any changes occurring in 2003, with variation in incentive strength arising from the proficiency target $y_{Tg,nclb}$ becoming relevant under NCLB.}

The difference $\hat{y}_{ijgs2003} - y_{Tg,nclb}$ captures how ‘far away’ each student is from reaching proficiency status without additional teacher effort. In terms of the model in Section 3.4, the predicted score $\hat{y}_{ijgst}$ may be written as the sum of all non-NCLB effort inputs in the production technology, $\hat{y}_{ijgst} = \gamma y_{ij}g'_{s't} + a_j + e_{st}^{abcs}$, representing the test score students would earn in 2003 had NCLB not been introduced. Taking the difference between realized and predicted scores in 2003 allows us to write the variable measured on the vertical axis in Figure 3.1 as a function of NCLB effort and measurement error in test scores: $y_{ijgst} - \hat{y}_{ijgst} = \epsilon_{ijgst} + e_{nclb}$. Since effort incentives are strongest for students who are predicted to score near the proficiency threshold ($\hat{y}_{ijgst} - y_{Tg,nclb} \approx 0$), these students should receive the most additional effort and gains over predicted scores should highest for them, exactly as shown in Figure 3.1.\footnote{To ensure that we do not systematically under- or over-predict test scores at certain parts of the distribution, we conduct the same exercise with a pre-reform year. As can be seen in Figure 3.1, we do a good job predicting test score outcomes in 2000 throughout the entire distribution, lending credence to the claim that the 2003...}
These non-parametric patterns are used to estimate a student-specific effort function that takes \( \hat{y}_{ijgst} - \hat{y}_{g,nclb}^T \) as its argument. We estimate the function by first differencing the year 2003 and 2000 profiles in Figure 3.1 and then fitting a tenth-order polynomial to the binned data using a weighted regression, with the weights capturing the number of students in each bin. The resulting effort function, denoted by \( e^{nclb}(\hat{y}_{ijgst} - \hat{y}_{g,nclb}^T) \), is plotted in Figure 3.7. The value of \( e^{nclb}(\hat{y}_{ijgst} - \hat{y}_{g,nclb}^T) \) is used as the student-specific value of effort each student receives in 2003.

**Identifying the Persistence of Effort**

With a measure of student-level effort in hand, we turn to estimating the rate at which this effort persists. In equation (3.13) of the model in Section 3.4, we posit that test scores in 2004 are determined in the following way:

\[
y_{ijgst} = \gamma \left( y_{ij'g-1's't-1} - e_{ij'g-1's't-1}^{nclb} \right) + a_j + e_{ijgst}^{abc} + e_{ijgst}^{nclb} + \gamma e_{ij'g-1's't-1} + \epsilon_{ijgst},
\]

where \( e_{s}^{abc} \) denotes the school-specific ABCs effort that results after NCLB’s disruption of ABCs targets. When modelling predictions about students’ likely performance without additional NCLB effort in 2004, we assume educators know the persistence rate of effort, \( \gamma^e \), forming the prediction as the sum of (i) the persistence of non-NCLB effort inputs, (ii) the effects of teacher ability and baseline ABCs effort, and (iii) the persistence of NCLB effort. Drawing a distinction between the component of the predicted score that is independent of the persistence of NCLB effort and that which depends on it, we let \( \bar{y}_{ijgst} \) denote the test score students would earn in 2004 had NCLB not been enacted in the prior year. In this case, there is no contemporaneous or persistent effect of NCLB effort and ABCs effort remains in steady state, implying that \( \bar{y}_{ijgst} \equiv \gamma \left( y_{ij'g-1's't-1} - e_{ij'g-1's't-1}^{nclb} \right) + a_j + e_{s}^{abc} \). Substituting \( \bar{y}_{ijgst} \) into the production technology, test scores in 2004 are expressed as the sum of predicted scores and patterns reflect student-specific NCLB effort.
contemporaneous effort responses:

\[ y_{ijgst} = \tilde{y}_{ijgst} + \gamma^e e_{ijg-1s't-1}^{nclb} + e_{ijgst}^{nclb} + \epsilon_{ijgst}, \] (3.23)

where \( \tilde{e}_{s}^{abcs} = \tilde{e}_{st}^{abcs} - e_{s}^{abcs} \) denotes the 2004 school-specific deviation from steady-state ABCs effort caused by NCLB’s disruption to ABCs targets.

When making decisions about how much NCLB-related effort to devote to each student in 2004, schools again take into account the distance between students’ predicted scores and the NCLB proficiency target, responding according to the NCLB effort function:

\[ e^{nclb}(\tilde{y}_{igst} + \gamma^e e_{ijg-1s't-1}^{nclb} - y_{g,nclb}^T). \] (3.24)

Equation (3.24) shows that the effort decision in 2004 depends on the effort students received in 2003 and the parameter \( \gamma^e \), thus highlighting the correlation of effort over time.

We account for this correlation with the following estimating equation for 2004 test scores, which is motivated by Equation (3.23):

\[ y_{igs2004} - \tilde{y}_{igs2004} = \gamma^e e^{nclb}(\tilde{y}_{g-1s'2003} - y_{g-1,nclb}^T) + \theta e^{nclb}(\tilde{y}_{igs2004} + \gamma^e e_{ig-1s'2003}^{nclb} - y_{g,nclb}^T) \]

\[ + \rho \tilde{e}_{s}^{abcs} + \epsilon_{igs2004}. \] (3.25)

The dependent variable is the difference between the realized test score and the score students would have earned in 2004 had NCLB not taken effect in 2003. By definition, it captures the effects of all new test score determinants that result from the incentives created by NCLB, which include the persistent effects of NCLB effort in 2003 and the contemporaneous effects of both NCLB and ABCs effort.

We define an empirical analogue to \( \tilde{y}_{ijgst} \) as the ‘counterfactual’ predicted score, which is constructed for students in grade \( g \) in 2004 by using our test-score prediction equation for that grade (estimated prior to NCLB) and substituting predicted grade \( g - 1 \) test scores from 2003 for realized 2003 test scores. Realized scores contain NCLB effort from 2003 and are
thus a function of $\gamma^e$. In contrast, since predicted scores capture what students would have scored in 2003 had NCLB not occurred, using them in the prediction equation ensures that the counterfactual predicted score represents the score students would earn had NCLB not taken effect in 2003.

Since the effort function and its inputs are both known in 2003, the effort devoted to each student in 2003 is observed and given by $e^{nclb}(\hat{y}_{ig-1'st-1} - y_{g-1,nclb}^T)$. In 2004, we let $e^{nclb}(\tilde{y}_{igs2004} + \gamma^e e^{nclb}_{ig-1'st-1} - y_{g,nclb}^T)$ represent the unknown value of effort and allow the parameter $\theta$ to amplify or diminish the NCLB effort response one year after the introduction of the reform. Since the input of the 2004 effort function depends on the persistence rate $\gamma^e$, we are unable to observe the level of effort received by students in 2004 directly without knowing the persistence rate. Yet in order to credibly identify the persistence rate, we need to account for the correlation of effort across time. We therefore use a maximum likelihood routine to jointly estimate $\gamma^e$, effort in 2004, $e^{nclb}(\tilde{y}_{igs2004} + \gamma^e e^{nclb}_{ig-1'st-1} - y_{g,nclb}^T)$, and the scale factor modifying the 2004 effort function, $\theta$.

Our routine also accounts for potential changes to ABCs effort incentives in 2004. The conceptual discussion in Section 3.4 showed that magnitude of NCLB’s disruption to ABCs effort is in large part a function of the difference between the persistence rates of non-effort and effort inputs and the average level of NCLB effort received by students in a given school in 2003, $(\gamma^a - \gamma^e)\sum_{g=3}^{G_s} e^{nclb}_{g-1'st-1}$. We use the student-level effort function $e^{nclb}(\hat{y}_{igt} - y_{g,nclb}^T)$ to calculate average school-level effort from 2003, which allows us to control for it directly in estimation routine in order hold constant the effects of ABCs incentives in 2004. Equation (3.25) is therefore modified by setting $\tilde{e}_{s2004}^{abc} = \tilde{e}_{s2003}^{nclb}$, with $\rho$ governing the effect of average effort (across all students) from 2003 on test scores in 2004.\footnote{We calculate $\tilde{e}_{s2003}^{nclb}$ as a jack-knife mean, leaving out the effort received by student $i$, to ensure that the estimates of $\gamma^e$ and $\rho$ are not confounded.}

We implement the maximum likelihood routine by assuming $\epsilon_{igs2004} \sim N(\mu, \sigma^2)$, treating $\mu$ and $\sigma^2$ as additional parameters to be estimated. The full parameter vector is given by $\omega = [\gamma^e, \theta, \rho, \mu, \sigma^2]'$ and the main parameters of interest, $\gamma^e, \theta, \rho$, are all separately identified. We establish separate identification by arguing that $\gamma^e$ and $\theta$ are separately identified, first
ignoring ABCs effort incentives. We then expand the argument to include ABCs incentives, showing that $\rho$ is separately identified from both $\gamma^e$ and $\theta$.

Identification Argument without ABCs Incentives

Conceptually, separate identification of $\gamma^e$ and $\theta$ requires that, conditional on 2003 effort, written $\epsilon_{nclb}(\hat{y}_{ig-1s'2003} - y_{g-1,nclb})$, there is remaining variation in 2004 effort, $\epsilon_{nclb}(\tilde{y}_{ig2004} + \hat{\gamma}^e\epsilon_{nclb}^{ig-1s'2003} - y_{g,nclb})$, and vice-versa. This condition is different from requiring that, conditional on the argument of the effort function in 2003, $\hat{y}_{ig-1s'2003} - y_{g-1,nclb}$, there is remaining variation in the argument of the effort function in 2004, $\tilde{y}_{ig2004} + \hat{\gamma}^e\epsilon_{nclb}^{ig-1s'2003} - y_{g,nclb}$. The latter condition is difficult to satisfy. The non-linear shape of the 2003 effort function ensures that the former condition is easily satisfied.

Suppose two students, $i$ and $j$, start 2003 with equivalent predicted scores, such that $\hat{y}_{ig-1s'2003} = \hat{y}_{jg-1s'2003}$. Since the NCLB test-score proficiency target does not vary across students, these two students have equivalent levels of NCLB incentive strength, $\hat{y}_{ig-1s'2003} - y_{g-1,nclb}$, and therefore have the same level of NCLB effort in 2003, $\epsilon_{nclb}(\hat{y}_{ig-1s'2003} - y_{g-1,nclb})$.

The two students also have very similar counterfactual predicted scores in 2004, $\tilde{y}_{ig2004}$ and $\tilde{y}_{jg2004}$, because the most important determinant of the counterfactual predicted score is the predicted score in 2003 (which is the same for both students). It is clear, then, that NCLB incentive strength is (nearly) perfectly correlated within-student over time, implying that there is no variation in 2004 incentive strength that is independent of 2003 incentive strength.

In a linear regression framework, it is therefore impossible separate the effects of 2003 and 2004 NCLB incentives on 2004 test scores.

It is important to note, however, that we do not estimate the separate effects of 2003 and 2004 NCLB incentives on 2004 test scores; instead, we estimate the separate effects of 2003 and 2004 NCLB incentives on 2004 test scores.

\[30\] Recall the counterfactual prediction is formed by replacing the realized prior-year math score with the prior-year predicted math score and then implementing the prediction equation on the prior predicted math score, prior reading score, and other covariates. Since the counterfactual predicted score also depends on these other covariates, the two students may have slightly different counterfactual predicted scores in 2004 if they have different reading outcomes in 2003 or different demographic variables in 2004.

\[31\] Indeed, looking ahead to the estimate we obtain for $\gamma^e$, the $R^2$ from a linear regression of $\tilde{y}_{ig2004} + \hat{\gamma}^e\epsilon_{nclb}^{ig-1s'2003} - y_{g,nclb}$ on $\hat{y}_{ig-1s'2003} - y_{g-1,nclb}$ is 0.97, implying that 2003 NCLB incentives explain nearly all of the variation in 2004 NCLB incentives.
2004 NCLB effort on 2004 test scores. We therefore require that, after conditioning on 2003 NCLB effort, there is remaining variation in 2004 effort. While 2003 and 2004 NCLB effort are correlated, one is not a perfect linear function of the other.

Intuitively, identifying variation comes from the non-linear shape of the effort function in 2003, which ensures that two students with the same level of effort in 2003 can have different levels of NCLB incentive strength and, correspondingly, different levels of 2004 NCLB effort. For example, refer to Figure 3.7 and consider two students who each have a 2003 effort level of 2, but one student has an incentive strength value of $-3$ while the other has a value of 11. Despite having the same level of effort in 2003, the students have different predicted scores in 2003 and continue to have different predicted scores in 2004. The student with prior incentive strength of $-3$ is still predicted to score relatively poorly but the extra effort he received last year moves him up in the incentive strength distribution, making him more marginal with respect to the test score target. This student receives more effort in 2004 than he did in 2003. The student with prior incentive strength of 11 also gets a bump in her predicted score, which also moves her up in the incentive strength distribution, but to a point where she receives less effort in 2004 because she is even farther away from the proficiency threshold and therefore less marginal.

Following similar logic, we use a more general argument to establish that, conditional on the 2003 effort function, identification of the parameters of interest requires a minimal assumption about the form of the 2004 effort function, namely that it should not be flat: we label this as the requirement that the function be “non-uniform.” Although we assume the same functional form for the effort function in 2004 as in 2003, this assumption is not required for identification. Indeed, any non-uniform effort function in 2004 is sufficient for identification.

To see this, suppose that the 2004 effort response is determined by some arbitrary non-uniform function and, as above, consider any two inframarginal students (one with a predicted score below the proficiency target and one above it) who receive the same level of effort in 2003, due to the non-monotonic nature of the effort function. Incentive strength in 2004 shifts rightward for each student by the common amount of 2003 effort that persists. A non-uniform effort function in 2004 then guarantees that at least some student pair satisfying the identical-effort condition in 2003 receives divergent levels of effort in 2004. Indeed, there is zero variation
in effort within all such student pairs only if the 2004 effort function is uniform, implying that any non-monotonic effort function in 2003 and non-uniform function in 2004 are sufficient for identification.

While separate identification of $\gamma^e$ and $\theta$ relies on the non-monotonic form of the 2003 effort function, note that the precise functional form is not assumed but, rather, estimated using our difference-in-differences strategy that exploits the introduction of NCLB in 2003 as an exogenous shock to incentives.

Identification Argument with ABCs Incentives

The key identifying assumption when adding the ABCs component is that ABCs incentives operate across schools while NCLB incentives operate within schools, thus providing separate identification of the incentive effects of the two schemes.

We rely on an important difference between the designs of NCLB and the ABCs to support this assumption. The ABCs sets only an average school-level growth target, implying that average school-level effort is critical in forming the likelihood of each school passing or failing the scheme. In contrast, NCLB sets a secondary, student-level target (the test score required for subject matter proficiency) in addition to its primary school-level target (the proficiency rate). The student-level target ensures that the distribution of student-level effort within a school is relevant in determining the likelihood of school-level success under NCLB.

The argument for separately identifying $\gamma^e$ from $\theta$ is the same as above, while the argument for separate identification of $\rho$ from both $\gamma^e$ and $\theta$ relies on the large within-school variation of NCLB incentives. As mentioned, given the fixed test score proficiency target of NCLB, a direct consequence of large within-school variation of student test scores is that NCLB incentives also vary widely across students within a given school. This ensures that there are a sufficient number of marginal and non-marginal students within all schools with differing values of $\bar{\epsilon}^{nclb}_{s2003}$, thus allowing us to separately identify $\rho$ from $\gamma^e$ and $\theta$. 
3.6.3 Results for Research Question Two

We apply the estimation routine to the sample of fourth grade students in 2004 who have non-missing math scores, third grade effort values from 2003, and counterfactual predicted scores; the resulting sample size is 86,237 students.

Table 3.5 presents the results. The first column provides an estimate of the persistence of effort without accounting for contemporaneous effort incentives. In this case, 50 percent of the initial effort effect persists one period into the future. As expected, this estimate overstates the true persistence rate, as shown by the estimates in column (2). Once we account for contemporaneous NCLB effort, the estimate of $\gamma^e$ falls to 0.13, implying that only 13 percent of the initial effort effect persists to affect 2004 test scores. The estimate of $\theta = 0.51$ implies that the effort response is scaled down by 50 percent in 2004 relative to 2003. The estimate of $\hat{\theta} < 1$ means that the difference between the effort received by marginal and non-marginal students becomes smaller in 2004 than in 2003 at the average school, suggesting a smaller redistribution of effort.

Accounting for ABCs incentives results in an estimate of $\hat{\rho} = 0.33$, which implies that a one standard deviation increase in school-level NCLB effort from 2003 results in a one percent of a standard deviation increase in student-level test scores. Although this is a relatively small effect, a positive and significant estimate of $\rho$ implies that NCLB effort responses from 2003 strengthened ABCs incentives for the average school in 2004, leading to student performance gains. Prior to NCLB, the average school passed the ABCs relatively easily, implying that most schools were non-marginal with respect to the ABCs and easily satisfied their targets. The NCLB responses in 2003 made it more difficult for schools to pass the ABCs in 2004, thus making the average school more marginal with respect to the ABCs and strengthening incentives.$^{32}$

It is important to note that the estimates of $\gamma^e$ and $\theta$ are nearly identical to those obtained from the maximum likelihood routine that does not account for ABCs incentives (the estimate of

$^{32}$This in the case for the average school. Some schools may find it difficult to pass the ABCs prior to NCLB, implying the NCLB response in 2003 may have made it even harder to pass in 2004. Such schools would become less marginal and reduce ABCs effort as a result.
\( \gamma^c \) is also more precise). This lends credence to the assumption underpinning the identification strategy: NCLB incentives vary within schools while ABCs incentives vary across schools.

### 3.7 Policy Analysis

Our results indicate that educators responded to the introduction of NCLB incentives in a meaningful way, exerting additional effort that affects student outcomes both contemporaneously and over time. These findings naturally raise the questions of whether there are gains to be had from implementing different incentive-based policies and how such policies compare to other proposals for improving average teacher productivity in terms of cost effectiveness.

We explore these questions in this section by examining the effects strengthening incentives under NCLB and comparing these potential reforms to other popular policies for improving average teacher performance. A key challenge in this exercise is that it is unclear how effort changes in response to financial incentives under a scheme like NCLB, as the program offers no explicit per-teacher financial cost or benefit when schools fail or succeed. Yet financial incentives provide a natural metric with which to compare the costs of different policies, as education agencies must be able to prepare thorough budgets for any proposed accountability reform.

Therefore, prior to conducting exploratory cost-benefit analyses, we first address this challenge by using our model and the connection between NCLB and ABCs incentives to determine the implied per-teacher financial cost of NCLB incentives. Our approach involves three steps. First, we calculate the degree to which schools lowered the probability of passing the ABCs by responding to NCLB, relative to the counterfactual scenario in which NCLB was not enacted. The differences in these passing probabilities and the ABCs bonus payment let us calculate the expected financial loss each school brought upon itself by responding to NCLB. Second, we use the expected financial loss along with our estimate of how schools respond to changes in ABCs targets from Section 3.6 (\( \hat{\rho} = 0.33 \)) to determine the relationship between financial incentives and educator effort. Third, we use this relationship and the observed NCLB effort response in 2003 to infer the implied monetary value educators place on the NCLB sanction.
Calculating Expected Financial Loss Under the ABCs

We begin by calculating the probability of passing the ABCs in 2004 for each school in the counterfactual scenario where NCLB was not enacted. Since the ABCs measure performance using contemporaneous test scores and set targets using prior scores, we require counterfactual test scores for each student from both 2003 and 2004. We use the predicted score in 2003 as the 2003 test score that would have occurred without NCLB and the counterfactual predicted score from 2004 as the test score that would have occurred in 2004. The differences between 2004 performance and the ABCs target (which uses lagged performance from 2003) for all students are then used to form school-level ABCs growth scores by following the aggregation rules under the program.33

We also calculate the probability of passing the ABCs in 2004, given that schools responded with additional effort in 2003. Of note, we do not incorporate 2004 NCLB and ABCs effort responses into the calculation, as we are interested in isolating the reduction in ABCs passing probabilities caused by the initial response to NCLB. In this case, we take the prior score for each student to be the actual prior score and the test score that would have occurred in 2004 to be the sum of the counterfactual predicted score and the persistence of the effort response from 2003. We again calculate the difference between 2004 performance and the ABCs target for each student and use these differences to form school-level ABCs growth scores.34

With school-specific ABCs growth scores, we are able to calculate the school-level probability of passing the ABCs, which is governed in our model by $F(\cdot)$, the cumulative density function of average measurement error in test scores at a given school. We represent $F(\cdot)$ using a normal distribution with mean zero and assess the sensitivity of our analysis to a variety of possibilities for the standard deviation of this distribution. Specifically, we let the standard deviation of school-level measurement error vary from 0.1 to 1 developmental scale points in

---

33While the model in Section 3.4 abstracts from some of the detailed rules for calculating school-level ABCs scores, this calculation follows those rules precisely. In particular, we sum the weighted and standardized grade-and-subject-specific average differences between realized and target growth within each school.

34For both the scenario in which NCLB never occurred and the one in which we examine NCLB’s impact on ABCs passing probabilities, we use realized reading scores as 2004 predicted reading outcomes in the calculations. ABCs reading targets depend on both prior math and reading scores, however, so despite using realized reading scores as 2004 outcomes in both scenarios, we do change the ABCs reading targets to incorporate prior counterfactual math scores where appropriate.
increments of 0.1.\footnote{For comparison, the standard deviation of the school-level ABCs score under the counterfactual scenario in which NCLB was not enacted is 0.34 developmental scale points.} In each case, we calculate the school-level probability of passing the ABCs assuming that NCLB was never enacted, the probability of passing the ABCs with only the 2003 NCLB response, and the difference between the two, reflecting the degree to which each school lowered the likelihood of passing because of their effort response to NCLB. For a 0.1 scale-point standard deviation of noise, the average difference between the two passing probabilities is 19 percentage points, while it is 8 percentage points for a 1 scale-point standard deviation of noise and monotonically decreasing in between.

The Relationship between School Effort and Expected Financial Loss

In Section 3.6, we showed that schools responded positively ($\hat{\rho} = 0.33$) to increased targets under the ABCs, as higher targets lower the likelihood of success under the scheme. We estimated that response by relating student-level test scores in 2004 to average school-level NCLB effort from 2003 – an object that determines the NCLB-induced change to school-level ABCs passing probabilities and thus implicitly determines the expected financial loss faced by each school. Using the NCLB-induced changes to ABCs passing probabilities calculated above, we now regress the change in the school-level likelihood of passing the ABCs in 2004 on average school-level effort in the prior year to obtain a direct relationship between effort in 2003 and expected financial loss in 2004. We then scale the estimate of $\hat{\rho} = 0.33$ by the resulting coefficient, thus obtaining an estimate of the direct relationship between expected financial losses and educators’ effort responses in 2004.

Regressing changes in ABCs passing probabilities on average school-level effort from 2003 results in a coefficient ranging from $-0.29$, when the standard deviation of noise is assumed to be 0.1, to $-0.08$, when the standard deviation is assumed to be 1. A one unit increase in average school-level effort in 2003 thus reduces the probability of passing the ABCs by a value between 8 and 29 percentage points. Multiplying these values by the ABCs bonus payment of $1500 per teacher and scaling the estimate of $\hat{\rho} = 0.33$ by the result implies that a $1 expected financial loss under the ABCs causes an effort-driven increase in average student test scores.
that is between 0.0007 and 0.0025 developmental scale points.

Calculating the Implied Valuation of the NCLB Sanction

We use these estimates to infer the value educators place on the NCLB sanction by relating the observed effort responses to NCLB to the estimated relationship between expected financial losses and effort. The average school-level NCLB effort gain in 2003 is 1.96 developmental scale points. Combining the average effort response under NCLB with the estimates above suggests that the NCLB sanction is valued between $784 (i.e 1.96/0.0025) and $2,800 (i.e. 1.96/0.0007) per-teacher.

Based on the observed NCLB response as well as the resulting variation in ABCs targets, and the ABCs financial bonus, we therefore estimate an upper-bound value for the NCLB sanction of $2,800.\textsuperscript{36} Note that this upper bound is likely overestimated to some degree, as it is derived under the assumption that chance plays an extremely large role in the probability of school-level ABCs success: the assumed standard deviation of school-level noise is 1 developmental scale point, three times larger than the observed standard deviation of the school-level growth score under the scheme.

A Cost-Effectiveness Comparison of Two Policies

The calculations above suggest that a program similar to NCLB that offers a bonus payment between $784 and $2,800 per-teacher when schools succeed could result in similar effort responses to those observed under NCLB. These figures help determine the approximate cost of such an incentive-based scheme, making it directly comparable to other policy reforms discussed by the prior work. In particular, Hanushek (2009 and 2011), Chetty \textit{et al.} (2014b), and Rothstein (2015) analyze policies designed to improve average teacher performance by dismissing teachers whose whose value-added falls in the bottom part of the measured distribution (for example,\textsuperscript{36}We implicitly assume that educators respond the same way to a given expected financial loss irrespective of whether the loss stems from the ABCs or NCLB. Conceptually, our first-order condition for optimal effort in equation (3.11) shows that this assumption rests on there being one effort choice governed by one cost function. Manipulating either the ABCs or NCLB object on the left-hand-side of equation (3.11) by a given amount then results in the same change to effort decisions.)
the bottom five percent).³⁷ We use our framework to conduct a cost-benefit analysis, exploring how an incentive-based reform compares to such ability-based policies.

We begin by placing the two policies on a common footing in terms of their effects on student achievement, before comparing their relative costs. Chetty et al. (2014b) provide benchmark estimates, indicating that replacing the lowest rated teachers with draws of new teachers would result in an average two standard deviation increase in ability for that subset. At the same time, based on estimates from Rothstein (2015), doing so would require a mean salary increase across all teachers of 1.4 percent as compensation for the increased employment risk. Although our sample differs somewhat, these estimates provide comparison values in our setting.

Our analysis suggests a potential avenue for raising teacher productivity by an equivalent two standard deviations, on average, through an incentive-based reform. The proposed reform involves setting tougher test score targets for currently non-marginal students and raising the value of the NCLB sanction (or reward under a comparable pecuniary scheme). In particular, the required two standard-deviation performance improvement can be attained by increasing the fraction of marginal students in each classroom from the current sample average, 26 percent, to 86 percent and raising the current value of the NCLB sanction.³⁸ The reduced-form estimates in Table 3.3 imply that a 60 percentage-point increase in the proportion of marginal students results in a performance improvement of 0.17 standard deviations of the test score, on average, which is equivalent to the impact of nearly a one standard deviation increase in teacher ability. The additional one standard-deviation performance improvement is achieved by raising the value of the NCLB sanction to 140 percent of its current value.³⁹ Such a policy results in a benefit that is comparable to replacing the bottom five percent of teachers (in terms of value-added performance) with average draws.

³⁷ The literature has also focused on reducing the attrition of the highest rated teachers. However, existing research (Chetty et al., 2014b) suggests that such a focus on the top is a less cost-effective (ability-oriented) reform than replacing the lowest rated teachers.

³⁸ Non-marginal students with predicted performance above the NCLB accountability target account for about 60 percent of our sample. It is therefore possible to increase the proportion of marginal students in each class by 60 percentage points (from an average of 26 percent) by setting more difficult test score targets for non-marginal students.

³⁹ To see this, note that a 60 percentage-point increase in the fraction of marginal students achieves a one standard-deviation ability increase in effort. An average proportion of marginal students of 86 percent thus achieves a 1.43 standard-deviation-of-ability effect (that is, \((0.86/60) \times 1 = 1.43\)), implying that a two standard-deviation performance gain is achieved by a sanction that is 140 percent \((2/1.43)\) of the current value.
The proposed reform serves as a thought experiment that depends on two assumptions. First, teachers are assumed to be able to continue to raise effort as progressively larger fractions of students become marginal – that is, the estimated linear relationship from Table 3.3 remains valid for larger values of the proportion of marginal students \( m_{jt} \) than are observed in our data.\(^{40}\) Second, we assume that increasing the value of the NCLB sanction also affects effort linearly, which requires that teachers’ observed effort choices are not associated with steeply rising marginal costs of effort. If teachers do face steep marginal costs, increases in the value of the sanction do not necessarily result in proportional increases in effort.

To compensate teachers for the new employment risk, implementing the ability-based reform requires a 1.4 percent increase in salary for all teachers, which amounts to an average increase of $700 per-teacher in North Carolina. Recall, the implied value of the NCLB sanction may be as high as $2,800 per-teacher and we need to further increase it to 140 percent of its current value under the proposed incentive reform. While it therefore appears that the incentive-based reform is unlikely to pass a cost-benefit test when compared to the dismissal-based reform, one must scale the required sanction value to account for the dismissal-based reform affecting only 5 percent of the teacher distribution. Thus, achieving a comparable average improvement with the incentive-based reform requires a sanction that is only 7 percent of the current value.\(^{41}\)

If the monetary equivalent of the NCLB sanction is less than $10,000 (i.e. $700/0.07), a sanction (or reward) that is 7 percent of the current value results in less than a $700 increase in per-teacher spending. Importantly, accounting for the differential long-run effects of effort and ability with our estimates of their respective persistence rates implies that the incentive-based reform is more cost-effective in the long run if the NCLB sanction is valued at less than $3,200.\(^{42}\)

The upper bound of $2,800 for the NCLB sanction falls below this threshold, suggesting that the

\(^{40}\)The binned-scatter plots in Figures 3.2 and 3.4 suggest that the relationship is indeed linear in every grade, although we cannot rule out deviations from linearity at more extreme values of \( m_{jt} \).

\(^{41}\)Since the ability-based reform affects only 5 percent of teachers, we scale 140 percent by 20 (140/20) to arrive at the 7 percent value.

\(^{42}\)The long-run comparison depends on how effort effects persist beyond one year – an issue we intend to investigate in future work. For that calculation, we note that Figure 3.6 shows that the effect of ability on test scores four periods into the future is 46.3 percent of the ability effect one period into the future (i.e. 0.19/0.41). Assuming a similar stabilization for the effects of effort, the effect four periods ahead amounts to 6 percent (i.e. 0.13\(\times\)0.463) of the initial effort effect. Thus, four periods forward, the incentive reform achieves 32 percent of the effect of the ability reform (i.e. 0.06/0.19). Scaling the short-run threshold by 32 percent (i.e. $10,000\times0.32) thus results in a long-run sanction threshold value of $3,200.
proposed incentive-based reform is potentially more cost-effective than the ability-based reform. Taking $2,800 as the true value of the NCLB sanction, 7 percent of the sanction amounts $196 per teacher. We scale this value to account for the effects of effort amounting to only 32 percent of the long-run effects of ability, which implies that the required per-teacher sanction amounts to $612.50, or 87.5 percent of the $700 required under the dismissal-based reform.

As a final comparison between the two policies, recall that the ability-based reform affects only a pre-determined fraction of teachers who fall in the bottom of the teacher performance distribution. Our calculations of the costs and benefits of the incentive-based reform are scaled to account for this feature of the ability-based reform, thus allowing us to compare the two policies directly. However, it is important to note that the incentive-based reform can be scaled to affect the effort decisions of the full distribution of teachers, while the ability-based reform is limited in this regard. Thus, the incentive-based reform potentially provides a viable, cost-effective policy lever for improving average teacher performance.

3.8 Conclusion

In this chapter, we have presented a two-part strategy permitting us, for the first time in the literature, to separate out teacher effort, which is responsive to education incentives, from teacher ability, which is not. This allows us to gauge the respective impacts of effort and ability on contemporary scores. Further, we measured the extent to which of these two potentially important education inputs persist differentially.

Central to our analysis was a novel identification strategy taking advantage of a natural experiment associated with the introduction of a federal accountability program in a setting – the state of North Carolina – where accountability incentives already operated. Specifically, we drew on the proficiency-count design of NCLB to construct a measure of incentive strength

43To see why, note that the ability-based reform entails firing the bottom 5 percent of teachers and replacing them with a random draw from the full distribution. In expectation, a newly drawn teacher’s value-added is equal to the mean value-added, implying that the marginal benefit of such a policy is declining as the dismissed teachers come from progressively higher positions in the value-added distribution. One would eventually face the prospect of dismissing teachers with above average value-added. At this point, the policy yields negative returns and forces a switch to policies geared toward retaining high-performing teachers, which the literature has already shown are inferior to policies based on dismissal.
for each teacher, showing a positive relationship between teacher value-added and this measure in the year NCLB was introduced but not in prior years. We exploited these differential relationships over time to separate teacher quality into teacher ability and the effort response associated with NCLB.

We found greater dispersion in teacher ability than in NCLB-induced teacher effort, but that both have significant effects on student achievement: a one standard deviation increase in teacher ability causes a 18 percent of a standard deviation increase in student achievement, while a one standard deviation change in effort leads to between a 3 and 6 percent of standard deviation change in achievement.

To evaluate the persistence of teacher ability and effort, we drew on a simple model of school-decision making to develop a structural estimation strategy, which allowed us to identify the persistence of effort separately from the effects of contemporaneous incentives. This evidence helped shed light on the extent to which ‘teaching to the test’ occurs in practice – a term often mentioned in the literature, but with little statistical evidence to indicate that it is important. Here, we found that effort has a significant positive effect on future test scores, but that the effect of effort persists at approximately 25 percent of the ability effect.

Based on our estimates and the model, we were able to explore the policy implications of rival education reforms. Our analysis indicates that using formal incentives constitutes a viable alternative means of accomplishing the goal of raising student and school performance, and is more cost-effective than competing ability-based reforms. Overall, our findings indicate that the effort margin is first-order: teacher effort is both a productive input and one that is responsive to incentive variation in a systematic way, with longer-term benefits for students.
Tables
### Table 3.1: Student-Level Summary Statistics

<table>
<thead>
<tr>
<th>Performance Measures</th>
<th>Mean 1</th>
<th>Std. Dev. 1</th>
<th>Obs. 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math Score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 3</td>
<td>144.67</td>
<td>10.67</td>
<td>905,912</td>
</tr>
<tr>
<td>Grade 4</td>
<td>153.66</td>
<td>9.78</td>
<td>891,971</td>
</tr>
<tr>
<td>Grade 5</td>
<td>159.84</td>
<td>9.38</td>
<td>888,469</td>
</tr>
<tr>
<td>Future Math Score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 6</td>
<td>167.16</td>
<td>11.01</td>
<td>739,386</td>
</tr>
<tr>
<td>Grade 7</td>
<td>172.61</td>
<td>10.70</td>
<td>617,669</td>
</tr>
<tr>
<td>Grade 8</td>
<td>175.79</td>
<td>11.36</td>
<td>503,091</td>
</tr>
<tr>
<td>Math Growth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 3</td>
<td>13.88</td>
<td>6.30</td>
<td>827,738</td>
</tr>
<tr>
<td>Grade 4</td>
<td>9.20</td>
<td>6.02</td>
<td>817,240</td>
</tr>
<tr>
<td>Grade 5</td>
<td>6.92</td>
<td>5.35</td>
<td>815,602</td>
</tr>
<tr>
<td>Reading Score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 3</td>
<td>147.03</td>
<td>9.33</td>
<td>901,235</td>
</tr>
<tr>
<td>Grade 4</td>
<td>150.65</td>
<td>9.18</td>
<td>887,153</td>
</tr>
<tr>
<td>Grade 5</td>
<td>155.79</td>
<td>8.11</td>
<td>883,689</td>
</tr>
<tr>
<td>Future Reading Score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 6</td>
<td>157.07</td>
<td>8.66</td>
<td>737,192</td>
</tr>
<tr>
<td>Grade 7</td>
<td>160.76</td>
<td>8.00</td>
<td>616,384</td>
</tr>
<tr>
<td>Grade 8</td>
<td>163.32</td>
<td>7.56</td>
<td>502,229</td>
</tr>
<tr>
<td>Reading Growth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 3</td>
<td>8.20</td>
<td>6.71</td>
<td>838,387</td>
</tr>
<tr>
<td>Grade 4</td>
<td>3.85</td>
<td>5.58</td>
<td>811,890</td>
</tr>
<tr>
<td>Grade 5</td>
<td>5.49</td>
<td>5.22</td>
<td>810,216</td>
</tr>
<tr>
<td>Demographics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College-Educated Parents</td>
<td>0.25</td>
<td>0.43</td>
<td>2,757,648</td>
</tr>
<tr>
<td>Male</td>
<td>0.51</td>
<td>0.50</td>
<td>2,778,454</td>
</tr>
<tr>
<td>Minority</td>
<td>0.40</td>
<td>0.49</td>
<td>2,776,729</td>
</tr>
<tr>
<td>Disabled</td>
<td>0.06</td>
<td>0.24</td>
<td>2,778,635</td>
</tr>
<tr>
<td>Limited English-Proficient</td>
<td>0.03</td>
<td>0.17</td>
<td>2,778,623</td>
</tr>
<tr>
<td>Repeating Grade</td>
<td>0.02</td>
<td>0.13</td>
<td>2,778,734</td>
</tr>
<tr>
<td>Free or Reduced-Price Lunch</td>
<td>0.44</td>
<td>0.50</td>
<td>1,998,653</td>
</tr>
</tbody>
</table>

**Notes:** Summary statistics are calculated for all third through fifth grade student-year observations from 1997 to 2005. The free or reduced price lunch eligibility variable is not available prior to 1999. Math scores are measured on different scales before and after 2001. We are able to convert second edition scale scores to their first edition counterparts for all tests except the grade three pre-test (the grade two test). Thus, all level and gain math score summary statistics are expressed on the first edition scale except grade three gains, which are calculated using first edition scores prior to 2001 and second edition scores for 2001 onwards. Future math and reading scores are the scores we observe for our sample of third to fifth grade students when they are in sixth, seventh, and eight grades. We use these scores when measuring the persistent effects of teacher ability and effort. We do not follow students past 2005, as the math scale changes again in 2006 but no table to convert scores back to the old scale is provided.
### Table 3.2: Teacher Performance Variables

<table>
<thead>
<tr>
<th>Teacher-Year Value-Added</th>
<th>Estimated Ability</th>
<th>Fraction of Marginal Students, $m_{ij}$</th>
<th>Estimated Effort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gr. 3</td>
<td>Gr. 4</td>
<td>Gr. 5</td>
<td>Gr. 3</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.17</td>
<td>-0.11</td>
<td>0.19</td>
</tr>
<tr>
<td>Observed SD</td>
<td>2.65</td>
<td>2.80</td>
<td>2.31</td>
</tr>
<tr>
<td>Estimated SD</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Obs.</td>
<td>24,105</td>
<td>22,246</td>
<td>20,596</td>
</tr>
</tbody>
</table>

**Notes:** This table presents means and standard deviations of the main teacher-level variables of interest. Summary statistics for teacher-year VA measures and for the fraction of marginal students in classrooms are calculated over all available teacher-year observations from 1997 to 2003. Due to the unavailability of second grade scores in 1996 and the change to the math developmental scale in 2001, we are unable to calculate marginal status for third graders in 1997 and 2001, and for fourth and fifth graders in 2002, thus explaining the smaller sample sizes. Summary statistics for estimated ability are calculated over all teacher-grade observations, where we include a teacher in a grade-specific distribution if she is ever observed teaching in that grade. A given teacher can be in more than one grade-specific distribution. The observed standard deviation is the raw standard deviation, while the estimated standard deviation is the estimate of the true standard deviation of teacher ability, obtained from the EB procedure. Summary statistics for estimated teacher effort are calculated across all teacher observations in 2003.
Table 3.3: The Effects of NCLB Incentives on Teacher Performance

<table>
<thead>
<tr>
<th></th>
<th>Third Grade</th>
<th>Fourth Grade</th>
<th>Fifth Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2003</td>
<td>2003</td>
<td>2003</td>
</tr>
<tr>
<td></td>
<td>Pre-NCLB</td>
<td>Pre-NCLB</td>
<td>Pre-NCLB</td>
</tr>
<tr>
<td>Effect of $m_{jt}$</td>
<td>1.55***</td>
<td>0.09</td>
<td>4.39***</td>
</tr>
<tr>
<td></td>
<td>(0.20)</td>
<td>(0.13)</td>
<td>(0.32)</td>
</tr>
<tr>
<td>Observations</td>
<td>2,144</td>
<td>10,452</td>
<td>2,598</td>
</tr>
</tbody>
</table>

Panel (b): Change in Teacher-Year VA as Dependent Variable

<table>
<thead>
<tr>
<th></th>
<th>Third Grade</th>
<th>Fourth Grade</th>
<th>Fifth Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2003</td>
<td>2000</td>
<td>2003</td>
</tr>
<tr>
<td>Effect of $m_{jt}$</td>
<td>2.08***</td>
<td>-0.40</td>
<td>4.11***</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.27)</td>
<td>(0.32)</td>
</tr>
<tr>
<td>Observations</td>
<td>2,651</td>
<td>2,393</td>
<td>2,453</td>
</tr>
</tbody>
</table>

Notes: In panel (a), we present estimates of $\rho$ from grade-specific regressions of equation (3.18). In the year 2003 regression, additional controls include teacher ability and teacher experience. The result in the pre-NCLB columns comes from a pooled regression of all pre-NCLB years that additionally includes year fixed effects. For third grade, the pre-NCLB years stretch from 1998 to 2000, and 2002; for fourth and fifth grade, they stretch from 1997 to 2001. In panel (b), we regress the change in teacher-year VA from 2002 to 2003 or from 1999 to 2000 on the fraction of marginal students in classrooms in 2003 and 2000, respectively, as well as cubic functions of 2002 and 1999 teacher-year VA. The reported coefficients are the effects of the fraction of marginal students within classrooms. Standard errors clustered at the school level appear in parentheses. *** denotes significance at the 1% level; ** denotes significance at the 5% level; and * denotes significance at the 10% level.
Table 3.4: Tests for Differential Sorting of Students to Teachers in 2003

<table>
<thead>
<tr>
<th></th>
<th>(1) Full Sample</th>
<th>(2) Third Grade</th>
<th>(3) Fourth Grade</th>
<th>(4) Fifth Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability</td>
<td>-0.0034***</td>
<td>-0.0010</td>
<td>-0.0046***</td>
<td>-0.0055***</td>
</tr>
<tr>
<td></td>
<td>(0.0008)</td>
<td>(0.0010)</td>
<td>(0.0011)</td>
<td>(0.0017)</td>
</tr>
<tr>
<td>$1(t = 2003) \times \text{Ability}$</td>
<td>-0.0033**</td>
<td>-0.005***</td>
<td>-0.0045**</td>
<td>0.0027</td>
</tr>
<tr>
<td></td>
<td>(0.0014)</td>
<td>(0.0019)</td>
<td>(0.0019)</td>
<td>(0.0025)</td>
</tr>
<tr>
<td>$N$</td>
<td>39,932</td>
<td>12,599</td>
<td>14,151</td>
<td>13,182</td>
</tr>
</tbody>
</table>

Notes: This table presents the results of regressions based on equation (3.21). The dependent variable in each column is the fraction of students in a teacher’s class who are marginal. Teacher ability is estimated using the EB estimator from equation (3.17), and we use the leave-year-out (or jack-knife) EB estimate in pre-NCLB years to avoid mechanical correlation between EB estimates and outcomes. Standard errors clustered at the school-level appear in parentheses. *** denotes significance at the 1% level; ** denotes significance at the 5% level; and * denotes significance at the 10% level.
Chapter 3. Education Production and Incentives

Table 3.5: Parameter Estimates

<table>
<thead>
<tr>
<th></th>
<th>(1) Without Contemporaneous NCLB and ABCs Incentives</th>
<th>(2) With Contemporaneous NCLB but Without ABCs Incentives</th>
<th>(3) With Contemporaneous NCLB and ABCs Incentives</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_e$</td>
<td>0.50$^{***}$</td>
<td>0.13$^{***}$</td>
<td>0.13$^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>$\vartheta$</td>
<td>-</td>
<td>0.51$^{***}$</td>
<td>0.47$^{****}$</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-</td>
<td>-</td>
<td>0.33$^{***}$</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>(0.06)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.97$^{***}$</td>
<td>0.87$^{***}$</td>
<td>0.30$^{**}$</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.12)</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>20.30$^{***}$</td>
<td>20.14$^{***}$</td>
<td>20.13$^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.14)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Observations</td>
<td>86,237</td>
<td>86,237</td>
<td>86,237</td>
</tr>
</tbody>
</table>

Notes: This table presents maximum likelihood estimates of variants of equation (3.25). The sample includes fourth grade students in 2004. The dependent variable in each column is the difference between the realized and counterfactual predicted math score. Standard errors calculated using the Outer-Product of Gradients method appear in parentheses. $^{***}$ denotes significance at the 1% level; $^{**}$ denotes significance at the 5% level; $^*$ denotes significance at the 10% level.
Figures

Notes: This is Figure 3 from Macartney et al. (2015). The figure is constructed as follows: In each year, we calculate a predicted score for each grade four student and then subtract off the known proficiency score target from this prediction – the horizontal axis measures the difference. We then group students into 2-point width bins on the horizontal axis. Within each bin, we calculate the average (across all students) of the difference between students’ realized and predicted scores. The circles represent these bin-specific averages: the solid circles represent year-2003 averages; the hollow circles are year-2000 averages. The figure also shows the associated 95 percent confidence intervals for each year. Standard errors are clustered at the school level.

Figure 3.1: Inverted-U Response to NCLB
Chapter 3. Education Production and Incentives

132

1.5 Teacher-Year VA

Proportion of Marginal Students in Class

(a) Third Grade

(b) Fourth Grade

(c) Fifth Grade

Notes: This figure depicts the relationship between teacher-year VA measures and the fraction of marginal students within a classroom in the year 2003. To construct the figure, we first group teacher-year observations into 20 equally-sized (vintiles) bins of the distribution of the fraction of marginal students on the horizontal axis. Within each bin, we calculate the average proportion of marginal students and the average teacher-year VA estimate. The dark circles in each panel represent these averages in 2003. The lines represent the associated linear fits, estimated using the underlying teacher-year data.

Figure 3.2: Teacher-Year Fixed Effects versus the Proportion of Marginal Students

Notes: This figure shows the distributions of teachers’ incentive-invariant abilities (which include base level effort). To construct the figures, we estimate equation (3.17), and construct EB estimates of teacher ability. Panel (a) shows the distribution of ability across all teachers. Panels (b), (c), and (d) show the distributions for teachers in third, fourth and fifth grades, respectively. We include a teacher in a grade-specific distribution if she is ever observed teaching in that grade. A given teacher can be in more than one grade-specific distribution.

Figure 3.3: Incentive-Invariant Ability Distributions
Notes: This figure illustrates teachers' 2003 effort responses. In panels (a) to (c), we present grade-specific partial relationships between teacher-year effects and the fraction of students in a teacher's class who were marginal. To construct these figures, we first residualize $m_{jt}$ with respect to the other controls in equation (3.18). For the pre-NCLB years, these controls also include year fixed effects. The horizontal axis measures residualized $m_{jt}$. We group teacher-year observations in 20 equal-sized groups (vinttiles) of the residualized $m_{jt}$ distribution on the horizontal axis. Within each bin, we calculate the average residualized $m_{jt}$ and the average teacher-year effect. The circles in each panel represent these averages. The lines represent the associated linear effects, estimated on the underlying teacher-year data. Panels (d) through (f) depict the relationship between the change in teachers' annual performance from 2002 to 2003 and from 1999 to 2000 and the fraction of students in their classes who were marginal in 2003 and 2000, respectively. To construct the panels, we first construct the change from $t - 1$ to $t$ between each teacher's teacher-year fixed effects from those years, as shown in equation (3.20). For each teacher-year, we then calculate the fraction of students in the 2000 or 2003 class who were marginal. We again group teacher-year observations into 20 equally-size (vinttiles) bins of the fraction marginal distribution on the horizontal axis. Within each bin, we calculate the average proportion of marginal students and the average change in teacher-year fixed effects. The circles in each panel represent these averages. The lines represent the associated linear fits, estimated on the underlying teacher-year data. In panels (g) to (i), we present grade-specific densities of 2003 effort levels. To construct these figures, we first obtain 2003 effort for each teacher by taking the linear prediction (fitted value) from equation (3.19). We then plot the distributions of these effort levels separately by grade.

Figure 3.4: Effort Predictions and Effort Distributions in 2003
Chapter 3. Education Production and Incentives

Figure 3.5: The Relationship Between Teacher Effort and Incentive-Invariant Ability

Notes: This figure illustrates the relationship between NCLB teacher effort, obtained as the fitted value from equation (3.19), and teacher incentive-invariant ability. We construct the figure by first grouping teachers into 20 equal-sized bins (vingtiles) of the ability distribution. Within each bin, we calculate average ability and average NCLB effort. The circles in each panel represent these averages. The lines represent the associated linear effects, estimated using the underlying teacher-year data.

Figure 3.6: Persistence of Teacher Ability and Baseline Effort in Pre-NCLB Period

Notes: This figure reports estimates of the $\beta_n$ coefficients from equation (3.22). Each estimate is obtained from a separate regression. The horizontal axis measures the number of year students are removed from their period-$t$ teacher while the vertical axis measures the impact of the period-$t$ teacher on students’ test scores in period $t+n$. The dark circles represent the estimated effects while the dashed lines represent the 95 percent confidence intervals with the associated standard errors clustered at the school level.
Notes: This is Figure 5(b) from Macartney et al. (2015). The horizontal axis is the same as in Figure 3.1. To construct this figure, we first take the bin-specific differences between the year 2003 and the year 2000 vertical-axis variable in Figure 3.1. The dark circles represent the resulting within-bin differences. We then estimate a tenth-order polynomial on the binned-data, weighting the regression by the number of student observations (across both 2003 and 2000) within each bin. At the extremes of the horizontal axis (less than -3 and greater than 22), we estimate linear regressions. The dark lines trace out the empirical effort function that results from these regressions.

Figure 3.7: Student-Specific Effort Function $e^{nclb}(y_{figst} - y_{9,nclb}^T)$
Bibliography


305-331.


Teachers: Teacher Value-Added and Student Outcomes in Adulthood.” NBER Working
Paper No. 17699. Cambridge, MA.

of Teachers I: Evaluating Bias in Teacher Value-Added Estimates.” *American Economic
Review*. 104(9): 2593-2632.

of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood.” *American

and the Assessment of Teacher Effectiveness” *Journal of Policy Analysis and Manage-
ment*, 23(2): 251-271.


