The Roles of Teacher and Peer Quality

on Human Capital Accumulation in Primary and Secondary School

BY

IRINA HOROI
B.A. University of Arizona, 2008
M.A. Central Michigan University, 2010

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Economics
in the Graduate College of the
University of Illinois at Chicago, 2016

Chicago, Illinois

Defense Committee:

Darren Lubotsky, Chair and Advisor
Steve Rivkin
Ben Ost
Javaeria Qureshi
George Karabatsos, Education Psychology

ACKNOWLEDGMENTS

I would like to thank my thesis committee - Darren Lubotsky, Steve Rivkin, Ben Ost, Javaeria Qureshi, and George Karabatsos - for their support and assistance. Without their guidance and advising I would not have been able to meet my research goals. An especially big thank you to my Chair, Darren Lubotsky, who kept me focused through the entire research process and to Steve Rivkin who elevated each of my contributions.

I would also like to thank my coauthors, Ben Ost and Moiz Bhai, for their assistance, ideas, and their efforts in our collaborations. Chapters 2 and 4 would not have been realized to the same caliber without their efforts.

Finally, I would like to thank the North Carolina Education Research Data Center for providing me with the data I used for this dissertation. Specifically, I would like to acknowledge Kara Bonneau who helped me put together my application, and has been most accommodating while I was working with the data.

CONTRIBUTIONS OF AUTHORS

Chapter 1 is an introduction chapter that touches on the overall theme of the thesis. Chapter 2 represents an unpublished manuscript for which I was the primary author. My cohort colleague Moiz Bhai assisted me with portions of the estimation strategy, specifically using the within family estimation techniques. In addition, he contributed to writing 25% of the manuscript. Chapter 3 represents an unpublished manuscript of my original work aimed at understanding if National Board certified teachers have lasting impacts on students' academic outcomes. Chapter 4 represents a published manuscript (Horoi, Irina and Ben Ost. "Disruptive Peers and the Estimation of Teacher Value-Added." *Economics of Education Review*, 49, (2015):180-192) for which I was the primary author. Dr. Ben Ost contributed to the scholarly idea and to writing parts of the manuscript – specifically the introduction and conclusion sections, in addition to general edits of the full text. Chapter 5 is my synthesis of the research presented in this dissertation, and my overarching conclusion including broad policy implications.

TABLE OF CONTENTS

CHAPTER PAGE

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF TABLES (continued)

LIST OF ABREVIATIONS

ED	Emotionally Disabled
IDEA	Individuals with Disabilities Education Act
NBPTS	National Board Professional Teacher Standards
NCERDC	North Carolina Education Research Data Center
SES	Socioeconomic Status

SUMMARY

Three separate studies were conducted in this thesis. Study number 1 (Chapter II) tests whether teacher quality measured by the National Board for Professional Teacher Standards certification (NBPTS) improves student achievement. Student-teacher matched longitudinal data that contains individual, school, and family identifiers on the universe of 4^{th}-8^{th} graders in North Carolina from 2007-2013 was employed. Empirical panel models were run that account for both permanent and time-varying differences between families, and for selection to classrooms and schools on both fixed and time-varying unobservables. Results show that students taught by NBPTS teachers have higher math and reading achievement. Analysis from stratification by grade suggests that the effect only exists for middle school teachers.

Study number 2 (Chapter III) tests whether the effect of NBPTS teachers persists one year out. Student-teacher matched longitudinal data that contains individual and school on the universe of 4^{th}-7^{th} graders in North Carolina from 2007-2013 was employed. Empirical panel models were run that account for unobserved fixed and time-varying differences between schools and school-grade combinations. Findings show that the effect of certified teachers persists for both reading and math. Relative to contemporaneous impacts, the effect depreciates only slightly for reading but by more than half for math.

The last study (Chapter IV) empirically tests how classroom disruption influences student learning and subsequently teacher evaluations. Student-teacher matched longitudinal data that contains student, teacher and school identifiers on the universe of 4^{th}-5^{th} graders in North Carolina from 2007-2012 was employed. Panel models were run that account for fixed and time-varying differences between schools. Findings show that students with serious behavioral difficulties substantially reduce the academic performance of their peers in 4th and 5th grades and average teacher value added for the grade.

I. INTRODUCTION

A. **The Importance of Teachers and Peers in the Production of Student Skill**

Policy makers and researchers alike are interested in factors that drive the production of student learning in schools. Their goal is to determining efficient policies to increase academic achievement and long-term outcomes such as college attendance and earnings. Teachers and peers are two inputs that researchers and policy makers have put a lot of emphasis on. This is driven by two considerations: one, classrooms are where the majority of the learning is generated through teachers and peer interactions, and two, the quality of classroom inputs matter for maximizing student learning. For example, Hanushek (2014) illustrates that good teachers, those who achieve one standard deviation higher in teacher value-added relative to the average teacher, could reduce the math achievement gap between student eligible for free-reduced lunch and those with higher socio economic standing by one quarter to one third of the average gap. Relative to interventions like reduction in class-size, identifying and retaining great teachers is nearly two to five times more effective at generating student gains depending on the grade and peer quality (Lazear 2001).

Peers effects are also important in the production of student skill and for education policy considerations. The efficacy of polices such as school choice, academic tracking, and local funding all rely on the non-importance of certain peer effects. For example, if peer effects are on average positive such that bad peers gain more from being with good peers than good peers loose from being with bade peers, schools with academic tracking will not maximize over all learning. Thus the extent and type of peer effects that affect student learning is incredibly important to generate policies that maximize learning. In addition, the existence of peer effects has implications for the identification of the effect of other classroom inputs on learning, as they can muddle the contributions of these inputs in ways that deter or wrongfully incite their importance.

In this thesis I focus on teacher and peer quality, and explore how distinct measures of these inputs map to student achievement both contemporaneously and in future periods. In chapter one and two I determine if teachers that are certified with the National Board Certification for Professional Teaching Standards (NBPTS) are more effective on average than non-certified teachers. The goal is to understand if this credential is a signal of teacher quality, which could be used to identify and reward quality teachers.

The last chapter estimates the effect of disruptive peers on own student achievement. Disruption is frequently reported as an issue by teachers and administrators, and thus is an important peer effect to understand. In addition, we consider the extent to which peer effects bias the estimated impact of other inputs by showing how students who are likely to be disruptive influence the estimation of teacher value added. While teachers are just one input whose estimated impact could be biased by peer effects, the use of value-added estimates in high-stakes personnel decisions makes it particularly important to correctly estimate teachers' impacts.

II. NEW EVIDENCE OF NATIONAL BOARD CERTIFICATION AS A SIGNAL OF TEACHER QUALITY

A. Introduction

No Child Left Behind and other recent education policies have increasingly focused on the quality of classroom instruction, paying particular attention to the role of teacher quality. In contrast to the consensus on the importance of teacher quality, there is much debate over the best way to achieve the goal, because it is difficult to identify and measure teacher quality. Schools tend to rely on internal evaluations by principals, external evaluations, or increasingly measures of teacher value added. Historically, public school principals have had the responsibility to hire, evaluate, and make tenure decisions, and evidence suggests that their ability to objectively evaluate teachers may be compromised from competing incentive structures from stakeholders, principal preferences, and costs faced from difficult interactions with teaching staff (Levy and Williams, 2004). For example Ho and Kane (2013) find an absence of differentiation in subjective teaching evaluations. One resolution is to identify teacher quality on the bases of test score value added. Alternatively, the use of outside raters also circumvents problems faced by subjective internal ratings, and it has the appeal of not relying upon the accuracy of value added measures, and potential adverse incentives such as teaching to the test, and whether the teacher is teaching a tested subject. However, external evaluations depend upon the assumption that raters get it correct.

In this paper we examine an increasingly important external teacher evaluation process, the National Board for Professional Teaching Standards (NBPTS) certification, and whether the certification is successful at identifying effective teachers. NBPTS is a voluntary certification process (teachers are only evaluated if they so choose). The organization examines applicants based on a rich portfolio meant to capture multiple dimensions of teacher quality (NBPTS, 2015). The portfolio includes a written component that allows teachers to demonstrate their teaching practice and evidence of their teaching ability with recorded videos of lesson plans and inclusion of teaching materials, in addition to a six part content knowledge assessment. Certification is awarded if the applicant's portfolio meets the standards set by the organization, and their assessment scores are above strict cutoffs. If these assessments and standards are correct, than certification will signal high teacher quality.

The key to our empirical evaluation of the effect of being taught by an NBPTS certified teachers on test-scores is the ability to address hurdles to the estimation of teacher effectiveness. Challenges introduced by unobserved heterogeneity and purposeful matching of students to teachers that results from preferences of families and decisions of administrators is likely to be a particular problem in this setting, because teachers have a label that signals quality to both administrators and families. Research on school quality has focused increasingly on time-varying student influences, and we adopt a set of approaches to account for matching of students to teachers on time invariant and time-varying shocks to student achievement. The first two empirical strategies account explicitly for time varying family shocks with the inclusion of a family-by-year fixed effect in a lagged achievement value-added model. Specifically comparisons are made between siblings in different schools, but in the same academic year. An analogous approach is to restrict the analysis to twins, refining comparisons to be between twin siblings assigned to different classrooms in the same academic year. This approach strengthens the family design, as twins share more commonalities relative to non-twin siblings.

Although the methods employing the family design provide potential improvements over traditional value-added models that use a combination of school and/or student fixed effects, there are still several potential deficiencies. First, if family shocks are not the primary time-varying changes that influence the sorting of children to classroom, the inclusion of family-by-year fixed effects will not solve the problem. Moreover, the focus on within family differences will tend to amplify the potential bias introduced by knowledge spillovers among siblings or family reallocation of education resources in response to observed teacher quality.

Consequently our preferred model uses grade level measures of the share of students with NBPTS teachers, and makes comparisons across cohorts within a school and year. Aggregation to the grade is particularly useful as it directly allows us to address sorting to classrooms on unobserved differences. With aggregation and school-by-year fixed effects only variation between cohorts is used to identify the effect of NBPTS on test scores.

Using longitudinal school administrative data from North Carolina of 3rd-8th grade students matched to teachers with sibling identifiers we find that NBPTS teachers outperform their non-certified peers. Our aggregate models reveal effects of about 0.04 of standard deviation in math and 0.015 of standard deviation in reading. The effects are similar to a reduction in class-size by two students in math, and one student in reading (Krueger and Whitmore, 2001; Krueger, 2003). Alternatively our disaggregate models that explicitly account for time varying family effects and time varying school effects are 25% smaller in math and similar in reading. The larger effects in aggregate models suggest that the direct effects to controlling for dynamic sorting to classrooms with family-by-year fixed effects appear to lead to underestimates of the certification effect. However, it is not clear if this is because of unobserved student heterogeneity or within family spillovers.

Comparisons of teacher performance before and after certification suggest that greater average effectiveness of certified teachers reflects fixed quality differences identified by the certification as opposed to human capital effects. Implementing policies with a primary goal to modify the effectiveness of teachers should place little weight on the NBPTS certification as a potential facilitator. Rather the certification can be used to reward more effective teachers where use of direct evidence on performance in the districts is not feasible.

Finally we explore heterogeneity by grade level. Our results show students of NBPTS teachers have larger achievement gains relative to non-certified teachers in middle school than in elementary, particularly in math. Such a result is consistent with stronger dependence on subject matter knowledge in middle school that can be assessed more accurately during the certification process. This is will particularly hold if the difficulty to acquire subject matter expertise is relatively higher for middle school grades.

The findings that NBPTS teachers are more effective than non- certified teachers may seem small. However, computing the present value of future earnings gains to students' using the estimated earning returns from a one standard deviation increase in teacher value-added, Chetty et al.[b] (2014), reveals that NBPTS teachers have substantial value; the present value of future earnings gains for the average class in our sample with a certified teachers equates to about $48,000.

B. <u>Conceptual Framework</u>

The origins of the National Boards for Professional Teaching Standards begin with the Carnegie Corporation report, "A Nation Prepare: Teachers for the 21st Century." The report identified teachers as key participants in rebuilding the education system and set out guidelines that defined what successful teachers should know and be able to do, in addition to supporting the creation of a rigorous assessment to see that certified teachers meet these standards. The NBPTS assessment today includes over 10 components that aim at evaluating the teachers ability in meeting 5 core standards: commitment to student learning; knowing subject matter and how to teach it; managing and monitoring student learning; systematically thinking about teaching practice and learning from experience; being a part of the learning community. Furthermore, before applying teachers must meet teacher experience and credential requirements. Certification is a voluntary process and can be costly in terms of time and money, however in many states and districts teachers receive pecuniary benefits from its receipt. The institutional facts lead directly to three theoretical sources that link NBPTS certification to teacher quality:

B.1 The Applicant Pool is Disproportionally High Quality Applicants

Teachers' perceived costs and benefits of applying for certification guide their decision to apply. Applying for NBPTS certification is costly because it requires 10 assessments, including subject knowledge testing and the creation of a teaching portfolio. In addition to the time commitment, the application fee is $2500. Successful certification is not guaranteed; the average passing rate is around 64 percent (Hakel, 2008). Since the certification period lasts for five years, with a renewal for additional five years before recertifying, the benefits to certification include higher monetary wages over the next ten years of teaching, plus the non-monetary benefits of having distinct value as a certified teacher. If these costs and benefits are correlated to teacher effectiveness, then the applicant pool will tend to be more (or perhaps less) skilled than the average teacher. Understanding the applicant pool is important for separating the effect of the certification program on the applicant pool from the Board's selection process itself.

To further illustrate the role of selection faced by the applicant pool we provide a selection model with two types of teachers, high quality and low quality, that face the costs and benefits given below (Spence, 1973). In this model potential teacher applicants maximize their expected monetary and non-monetary benefits given the costs, and as a result the decision to certify depends on whether or not expected lifetime benefits are higher than the costs. Given that the teacher certification process attempts to assess effectiveness through complex assessments, high quality teachers face a lower certification cost given by C_H because they require less time for preparation than lower quality teachers who face a certification cost given by C_L. Furthermore, schools prefer certified teachers, because they see certified teachers as higher quality, and therefore provide a larger payout for certification[1].

Given that wages are higher for certified teachers at W_c than wages for non-certified teachers, W_{NC}, and given that the expected benefit, which is a function of the wages W_c, W_{NC} and the non-monetary benefit denoted by ω, will be higher for the high quality teachers, two possible outcomes can be derived. The first scenario produces self-selection of only high quality teachers into the applicant pool. The first outcome holds if the expected benefit minus costs is greater than zero only for high quality teachers:

$$E^H(W_C, W_{NC}, \omega) - C_H > 0$$

$$and \;\; E^L(W_C, W_{NC}, \omega) - C_L < 0$$

Alternatively, if expected benefits for both high quality and low quality teachers are higher than the costs they incur, both high quality and low quality teachers would apply.

$$E^H(W_C, W_{NC}, \omega) - C_H > 0$$

$$E^L(W_C, W_{NC}, \omega) - C_L > 0$$

A modified version of this model that likely depicts the reality better is if certification costs and benefits faced by teachers are continuous and a decreasing function of teacher experience. If costs for high quality teachers are lower at every experience level and the expected lifetime benefits are the same for both quality teachers, then a larger proportion of high quality teachers would enter the certification applicant pool.

[1] Over 30 states and district have salary increases or bonuses for holding NBPTS certification. In North Carolina, the state for which our data represents, gives teachers a 12% increase in their pay from holding this credential (National Board for Professional Teaching Standards Certification: What Legislators need to know).

Although data limitations preclude us from identifying the applicant pool, we describe teacher, school, and career characteristics of certified teachers and non-certified teachers in Tables I and II to infer about the likely applicant pool. We observe certified teachers on average gain certification by the time they have 11.36 years of teaching experience, suggesting that mid-career teachers face lower costs as our selection model proposed (Table II).[2] About 45% of NBPTS teachers have an advanced degree when they are certified and 55% have bachelors. Surprisingly, NBPTS certified teachers tend to teach at marginally poorer and with larger proportion of Hispanic students after they certify (Table II). Compared to non-certified teachers, certified teachers achieve 30% of a standard deviation higher on their Praxis exams, and are 21 percentage points more likely to have an advanced degree. Although the descriptive statistics cannot reveal whether the ability difference is driven by the quality of the applicant pool or the quality of the NBPTS evaluation process, they do provide auxiliary evidence indicating that NBPTS teachers are higher quality relative to non-certified teachers.

B.2 The NBPTS Evaluation Standards Correctly Identifies High Quality Applicants

The NBPTS rigorously evaluates teachers using exams to test subject knowledge expertise, and rating standards developed by a team of professional educators to grade the required teaching portfolios. If the evaluation standards accurately measure teacher effectiveness in the classroom, then the receipt of certification is a direct link to observed teacher quality. Although we are unable to explicitly test the claim that NBPTS application selects high quality teachers, because we do not observe the full applicant pool, the two studies that do observe the full applicant pool reach mixed conclusions. First, Cantrell et al. (2008) report that teachers that fail to achieve NBPTS certification are less effective at improving academic achievement than teachers that never apply, however they do not find statistically significant differences on student

[2] The NBPTS requires that teachers have at least a minimum of 3 years of teaching experience in order to apply.

achievement between certified teachers and those who never applied.[3] Goldhaber and Anthony (2007), on the other hand, find certified teachers to be no more effective at improving reading achievement and are more effective in math than non-certified applicants.[4]

B.3 The Certification Process Increases Teachers' Human Capital

Completing the application process provides several opportunities for learning. First, potential applicants are required to take a multi-part subject knowledge test, and preparation for this assessment may improve applicants' understanding of subject specific knowledge. Next, the portfolio entries require lesson plans and reflections on their effectiveness, which may illuminate teachers on their weaknesses enabling improvements. A number of studies testing for teaching capital effects due to NBPTS certification find that the effect of NBPTS teachers on academic achievement remains unchanged post certification (Harris and Sass 2009; Clotfelter et al. 2007; Goldhaber and Anthony 2007; Chingos and Peterson, 2011).

C. Identification Strategy

A paucity of research has examined the impact of National Board for Professional Teaching Standards (NBPTS) certification on student achievement, and found effects of certification on test scores in the range of 0.0 to 0.05 of a standard deviation of the standardized test-score gains distribution (Anthony and Goldhaber, 2007; Clotelter et al., 2007; Harris and Sass, 2009, Cantrell et al., 2008; Cowan and Goldhaber, 2015). The work on NBPTS certification attempts to address concerns on sorting through the use of value added models with school fixed effects (Anthony and Goldhaber, 2007, Clotfelter et al. 2007), student fixed effects (Harris and Sass, 2009), using schools where sorting is balanced on observables (Clotfelter,

[3] While this study uses experimental data, it is should be noted that the experiment was likely imperfect because the comparison group of non-certified teachers were chosen by the principle of the teacher's school and not randomly.
[4] This study also uses administrative data from the North Carolina Department of Public Instruction for academic years 97-99, during which time the NBPTS utilized different standards. The second wave of standards began in 2002, and are currently in place.

Ladd, and Vigdor, 2007; Goldhaber and Cowan, 2015), and tracking fixed effects (Goldhaber and Cowan, 2015). The sole study employing an experimental design, albeit imperfect, finds positive effects, however, the estimates are imprecise (Cantrell et. al, 2008).

Despite the attempts to address sorting through the use of test score value added models with student, school, or tract fixed effects, many of the studies on NBPTS certification only address within school sorting on limited observables, with few accounting for unobserved ability in a time invariant manner. Emerging evidence (Rothstein, 2010) shows the complexity of sorting may not be limited to levels of achievement, but it may also occur on other measures such as on achievement gains, which creates challenges in evaluating the role of teachers on academic achievement. Horoi and Ost (2015) provide suggestive evidence of sorting on non-cognitive attributes such as emotional disabilities. Sorting on non-cognitive measures poses a larger problem, because it is likely unobserved to the econometrician. Moreover non-cognitive measures are not only correlated with students' own achievement but also their peers' achievement. Thus their absence in models has potentially large ramifications.

A final concern arises from contemporaneous shocks experienced by students either from their home, neighborhood, or school environments. For example students may experience a family member losing their job, which may have multiple causal linkages to poorer academic outcomes. If students predisposed to these situations are systemically sorted to teachers, then estimates of certification would be biased by the family shock. Many studies address school-shocks in a good manner by using across classroom variation within a school and year (Clotfelter et al., 2007; Cowan and Goldhaber, 2015), however no study yet has been able to address neither family nor neighborhood shocks.

To overcome the potential selection concerns we use several different empirical strategies to isolate variation in exposure to NBPTS certified teachers. First, we use lagged achievement value-added models employing with-in family variation. Specifically we compare siblings in the same academic year but in different schools by employing sibling-by-year and school-by-year fixed effects. Lagged achievement is widely used to deal with unobserved heterogeneity, however, it may be an imperfect measure of ability and thus insufficient. The availability of family identifiers allows us to go beyond traditional models by using variation within family-by-year. The family-by-year models address unobserved fixed and time-varying family differences. Furthermore, siblings are relatively more similar to one another and share a portion of the same genetic make-up, thus these models mitigate concerns over unobserved heterogeneity relative to student comparisons across classrooms.

Despite the potential advantages of these models, they do face some limitations. Similar to traditional models the concern of non-random placement of students to teachers remains a possibility. Unobserved differences among siblings may be related to classroom placement, and given that sibling-by-year and school-by-year fixed effects limit the variation used to identify the certification effect, this could introduce substantial bias. Knowledge spillovers among siblings is an additional concern, because they can introduce a downward bias. A final concern pertains to the potential dynamic response of parents to differences in their children's academic achievement. If for example parents allocate more resources to their poorer achieving children and positive sorting on ability is also a concern then this may underestimate the effect on NBPTS certification.[5] On the other hand, if parents respond by allocating more resources to the higher achieving student this would exacerbate the upward bias driven by the positive sorting. Unfortunately we are unable to test for parental responses. However, the literature on intra-

[5] We specifically write may lead to an underestimate, because it depends on the extent of the positive sorting.

family resource allocation is mixed. Whereas some studies find that parents act by reinforcing differences (Frjjter et al., 2013; Rosenzwieg and Shultz, 1982) others find that parents compensate for the inequality (Berhamn et al. 1982) or that they do neither (Royer, 2009; Kelly, 2011).

Our data also identifies twin pairs, which allows us to estimate models using within family variation where we can compare twins in the same academic year. Comparisons within twins offer an improvement over both student comparisons and sibling comparisons, because twins share the same age, many of the same environments at the same developmental stages, and genetic make-up, and therefore their comparisons reduce unobserved differences. Despite the improvements in abating bias, the same potential problems that are a concern for within sibling comparisons are also a concern for within twin comparisons.

To assess the degree to which within sibling sorting and sibling spillovers are problematic we run several tests. To understand the extent of classroom sorting we run linear probability models where we predict the probability being taught by a NBPTS certified teacher with observable student characteristics using three sources of variation: within school-by-year, within siblings-by-year, and twins-by-year.[6] For the within-family models we restrict observables to the set that would vary within siblings.

Since the within family models possibly suffer from some limitations, we estimate preferred models using across cohort comparisons within the same school. Specifically we aggregate our treatment variable, whether the student has a NBPTS teacher, to the grade and regress the share of students in the grade taught by NBPTS teachers on test scores. Using lagged achievement value-added models with school-by-year fixed effects, we isolate the effect of

[6] Student observables include: lagged test-scores in both subjects, age, and indicators for low SES, genders, race, behavioral disability, other disability, and limited English proficiency.

having a NBPTS teacher on student achievement, by looking at differences in achievement across grades in the same school and year due to differences in the share of students in the grade taught by NBPTS teachers.

Unlike within family models, using variation across cohorts addresses concerns of student selection. Although it is likely that students get sorted to teachers based on observable and unobservable attributes, systemic selection to grade is unlikely. In addition, by including a school-by-year fixed effect we address the concerns on sorting to schools on fixed and time-varying attributes. One possible validity concern is if differences in cohorts across grades in the same school and year are related to differences in the proportion of NBPTS teachers in those grades. Such a concern however is highly unlikely because studies have demonstrated that switching teachers to teach different grade negatively affects student outcomes (Ost and Schiman, 2015). Nonetheless, as a further robustness check we evaluate whether observable grade level characteristics predict the proportion of student in the grade with NPBTS teachers using the same cohort variation as in our preferred model.

D. <u>Empirical Models</u>

D.1 Classroom Level – Within Family Variation

To estimate the impact of NBPTS certification on student achievement we estimate a lagged achievement value-added model with an indicator for whether a student was taught by an NBPTS certified teacher.

$$(1)\ A_{ifcgst} = f\left(A_{ifcgst-1}\right)\lambda + \beta NBPTS_{ct} + X_{ift}\,\delta + \bar{X}_c\pi + T_c\rho + \phi_{ft} + \theta_{st}\ + \epsilon_{ifcsgt}$$

To identify the effect off the variation from siblings in different schools we include sibling-by-year, ϕ_{ft}, and school-by-year fixed effects, θ_{st}. We control for lagged achievement using a cubic expansions in prior test scores in both math and reading. Additionally we include a vector of student characteristics, X_{ift} , age, birth order, spacing of siblings, and indicators for race, gender, disability, limited English proficiency indicator, and economically disadvantaged; a vector of classroom characteristics, $\bar{X}_c$, mean subject-specific lagged test-score and age, class-size, and proportion non-white, limited English proficient, disabled and economically disadvantaged; vector of teacher characteristics, T_c, experience dummies and an indicator for advanced degree. We also include grade-by-year dummies to account for changes in curriculum and tests.

Equation (1) is run separately for math and reading achievement, and all standard errors are clustered to the teacher-by-year level. For comparisons to models often run in the literature, we also estimate models with just school-by-year, and school-by-grade-by-year fixed effects. The models are estimated on a sample of siblings where in each year at least two siblings are in 4th-8th grades.[7] We also estimate models where we restrict the sample to only twins thereby making comparisons within twins. In this case we estimate equation (1) where we exclude the school-by-year and grade-by-year fixed effects since twins are in the same school, grade, and year, and many of the student characteristics that do not vary between twins such as age and race. Furthermore, the sibling-by-year fixed effects are replaced by twin-by-year fixed effect.[8] For a similar comparison to the sibling sample we estimate model (1) replacing school-by-year fixed effects to school-by-year-by-grade fixed effects[9]. Similar to comparisons within twins, this

[7] To provide evidence that these estimates are generalizable, we estimate (1) on the full sample of students with just school-by-year fixed effects, and another specification with school-by-year and school-by-grade fixed effects. Results are nearly identical to the same specifications with the sibling sample.
[8] We also drop birth order and sibling spacing.
[9] We also exclude grade-by-year fixed effect as there is no variation in these dummies within a school-grade-year.

specification accounts for unobservable differences between cohorts, in addition to accounting fixed and time varying differences between schools, neighborhoods, and families.

D.2 Grade Level – Cohort Variation

Limitations of using within family variation at the classroom level can be remedied through another set of models, which employ across cohort variation within a school and year. Specifically we estimate lagged achievement value added models with school-by-year fixed effects, where the estimate of interest is on the proportion of students in the grade taught by NBPTS teachers. A variable that varies at multiple levels can be split into between and within variables that are mechanically unrelated to one another, thus aggregating our variable of interest to the grade level eliminates the problematic classroom level variation (Rivkin et al., 2005).

$$(2)\ \ A_{ifcgst} = f\left(A_{ifcgst-1}\right)\lambda + \beta\overline{NBPTS}_{gst} + X_{ift}\,\delta + \bar{X}_{gst}\pi + \bar{T}_{gst}\rho + \theta_{st}\ + \epsilon_{ifcsgt}$$

Similar to the classroom level models we include lagged achievement using cubic expansions in prior test scores in both math and reading, the exact same student controls and grade-by-year fixed effects. The models differ, however, as our variable of interest, $\overline{NBPTS}_{gst}$, and other teacher credentials, $\bar{T}_{gst}$, are aggregated to the grade such that each student-by-year observation receives the grade level mean of the variable in question. Instead of including classroom characteristics, we include grade characteristics of the same variables, $\bar{X}_{gst}$. We estimate these models using the sibling sample described in section A, and all standard errors are clustered at the school-by-grade-by-year level[10,11] In addition, we estimate models where we include both

[10] This is done for comparison purposes.

school-by-year and sibling-by-year fixed effects for comparison purposes with the same fixed effect specification at the classroom level as these models test for knowledge spillovers.

E. Data

This study uses school administrative data with matched teachers to student records from the North Carolina public schools housed at the North Carolina Education Research Data Center (NCERDC) for grades 3 to 8 from 2006 to 2013. Student variables include race, gender, economically disadvantaged status, limited English proficiency status, disability status, age and end-of-grade standardized test-scores in both math and reading. Teacher characteristics are pulled from teacher pay records, and include years of experience, educational attainment, and national board certification status. Since we use lagged achievement models, and standardized testing does not begin until students reach third grade in North Carolina, we use third grade achievement as the baseline measure for lagged achievement for the students in 4^{th} grade. In addition, our estimation sample begins with the cohort from 2007, we use lagged achievement from 2006 as the baseline achievement for 2007.

To match students to their subject specific teacher and peers, we use course-membership files to group students on year, school, course title, semester and section. This procedure identifies the students' subject specific classroom. We restrict the analysis to math and reading classroom(s) and run models separately by subject. Finally, using data from the North Carolina Center for Health Statistics, we match students born in the state of North Carolina from 1987-2009 to their siblings born from the same mother as long as they are enrolled in a public school through the study's time period.

[11] To provide evidence that these estimates are generalizable, we run (2) on the full sample of students with just school-by-year fixed effects. Results are nearly identical to the same specifications on the sibling sample and can be found in appendix A, Table 1.

In Table III we show descriptive statistics for the sibling, twin, and full student samples. Examining the sibling and full samples reveals that both samples are similar on key classroom attributes such as class size, NBPTS certification, teacher experience, and teacher education. However, additional comparisons on the composition of the sibling and full sample reveals divergent of similarities on measures of socioeconomic and demographic outcomes. The divergence in the composition of the siblings and full samples suggests that our estimates are potentially less generalizable for the wider population. A similar observation can be made by comparing twins to the full student sample.

A final note concerns measurement error: the NCERDC administrative data includes the entire population of students attending public schools, which improves precision over survey data. Nevertheless, measurement error concerns might still arise because of how we classify teacher to student matches. Students may begin the school year with one teacher but may switch to another teacher during the semester, and such switches may reintroduce minor measurement error in the data pertinent for classroom level models. Consequently, our grade-level models do not face this issue.

F. <u>Results</u>

F.1 Main Results – Classroom level

In Table IV we explore the effects of having an NBPTS certified teacher on achievement in math and reading. Each cell reports estimates from a separate model. Focusing on the results for math in Panel A, we find that our base specification in column one, a traditional lagged achievement value added model that accounts for a rich set of covariates, produces a statistically significant effect of having an NBPTS certified teacher of 0.047 of a standard deviation on average. The addition of school-by-year fixed effects in column two reduces the effect to 0.036

of a standard deviation, nearly a 50% reduction due to accounting for school sorting and contemporaneous school shocks. The inclusion of sibling-by-year fixed effects to school-by-year fixed effects in column three, reduces the effect by only 11%, however it remains statistically indistinguishable from the coefficient in column two. The similar findings in columns two and three illustrate that either lagged achievement and the included variables capture within school sorting, or the siblings-by-year fixed effects do not capture the salient differences that lead to sorting. In columns four and five we examine the sensitivity of our results to the alternate controls for differences in schools. Examining the differences between coefficients in columns two and three and between four and five, where we replace school-by-year fixed effects with school-by-grade-by-year fixed effects shows that the NBPTS coefficients are not statistically different from one another.

Panel B. in Table IV presents the results for NBPTS certification on achievement in reading. The effectiveness of NBPTS teachers is considerably smaller for the full sibling sample in reading relative to math, as the size of the NBPTS coefficient varies from 0.013-0.019 of a standard deviation for the former and 0.027-0.047 of a standard deviation for the latter. Our base specification in column one shows that students of certified teachers outperform students of non-certified teachers on average by 0.017 of a standard deviation. Accounting for school-by-year fixed effects in column two reduces the effect to 0.013 of standard deviation, which is statistically indistinguishable from one at the 5% tolerance level. By comparing columns two and three we find that accounting for permanent and time-varying differences among families does not change the size or significance of the reading coefficients. Additionally, we find that as we improve in our ability to account for unobserved differences between cohorts in columns four and five our effect sizes do not vary relative to column two and are only slightly larger to column three.[12]

In Table V we examine the sensitivity of the NBPTS coefficient when the sibling sample is restricted to only twin siblings. Beginning with Panel A. column one which runs a slight variation of the base specification in Table IV, we find the effect of a NBPTS teacher on math achievement to be about 0.04 standard deviations. Column two includes school-by-grade-by-year fixed effects and we find an insignificant effect of .015 standard deviations. In column three we replace these fixed effects with twins-by-year fixed effect that capture more of the heterogeneity among students. We find that twins assigned to an NBPTS teacher on average have higher achievement by 0.029 of a standard deviation in math than their twin sibling with a non-certified teacher. The result is 100 % larger than comparisons made within schools presented in column two, however these estimates are statistically indistinguishable from one another. Sample variation and heterogeneous treatment effects likely drive estimate differences in column two from the full twin sample. The similar estimates across siblings more generally likely reflects that twin sets that face different teachers are similar to siblings who are close in age but in different grades.

Panel B. of Table V shows the reading results. In our base specification we find that an NBPTS teacher raises achievement by 0.027 standard deviations. In columns two we add school-by-grade-by-year fixed effects and find an effect of 0.023 standard deviations. Nevertheless, when we exchange school-by-grade-by-year with twin-by-year fixed effects in column three the coefficients are similar in size at 0.025 standard deviations. By comparing across columns, we infer either the existence of negative selection to NBPTS reading teacher, or there exists

[12] Column two should be compared to column four and column three should be compared to column five.

treatment heterogeneity between twins and other siblings, because the effects are substantially larger in the twin sample.

F.2 Testing for Student Sorting

In tables IV and V we show that the NBPTS coefficients are insensitive to permanent and time-varying family unobservable characteristics. We further investigate the association between pre-observed student characteristics and the probability of having a NBPTS teacher for math both within schools more generally and within families in Table V1. [13, 14] The first column assesses student sorting within schools by using a linear probability model with school-by-year fixed effects. Columns two and three test for within family sorting by running linear probability models with siblings-by-year fixed effects in the former and twins-by-year fixed effects in the latter. The last column uses across cohort variation within a school-by-year.[15] Notably in column one we do observe that a number of student attributes are associated with placement with an NBPTS teachers. For example, a low SES student is 1.1 percentage points less likely to have an NBPTS teacher, whereas a student who achieves one standard deviation higher on his or her lagged math achievement test is 1.7 percentage points more likely to have a NBPTS teacher. We also see some evidence of sorting on non-cognitive attributes. For example, we see sorting on the category of other disability, as students with other disabilities have an increased association of having an NBPTS certified teacher by 0.4 percentage points. Although the associations remain small, it is consistent with negative student selection to NBPTS teachers. These statistically significant associations on both cognitive and non-cognitive attributes also raise the possibility that sorting may also occur on other unobserved attributes.

[13] Pre-observed student characteristics refer to characteristics observed with a one year lag.
[14] Other observables included in these models, which are not shown include race indicators, age and grade-by-year fixed effects. These controls are not included in column 3.
[15] Samples are constrained to school years with both certified and non-certified teachers. Column 3 is also constrained to only twins.

Furthermore, columns two and three show that even within families the lagged math achievement is positively associated with the probability of taking math with a NBPTS certified teacher. The within twin-by-year model shows the largest effects of those in columns 1-3. Column three also shows that twins with a disability other than a behavioral disability are 1.9 percentage points less likely than their non-disabled twin to have a NBPTS certified teacher for mathematics. In contrast, in column four we observe only one marginally significant relationship between the share of students in the grade taught by NBPTS teachers and the share of student in the grade that are black.

F.3 Main Results – Grade Level

In order to address threats of within school sorting that classroom level models suffer from we aggregate the proportion of NBPTS teachers to the grade level and estimate equation (2) and similar variations. Panel A. of Table VII contains the results for math achievement and Panel B. contains the results for reading achievement. The effects for the NBPTS coefficient at the grade level can be interpreted as the change in student average achievement that would occur if all students in the grade had NBPTS teachers versus none of the students having an NBPTS teacher. Our base specification with no fixed effects in column one shows that the effect of being in a grade with all certified teachers on average improves math score by 0.065 of a standard achievement and reading score by 0.02 of a standard deviation. In column two we add school-by-year fixed effects identifying or effect from across cohort variation within a school and year. By accounting for school sorting and contemporaneous school shocks we find that effect reduces to 0.041 of standard deviation in math and 0.012 of standard deviation in reading. In column three we add sibling-by-year fixed effects, and we find that the effect on NBPTS remains unchanged for math and is slightly smaller and statistically indistinguishable from zero for reading. We infer from the minor changes in the coefficients that family unobservables are unrelated to grade level differences. The grade level specifications such as the classroom level specifications are consistent in magnitude. On the other hand, the grade level findings are approximately 15-30% larger for math than the comparable regression using the classroom level measure.

Overall, the results from tables VI - VII provide evidence that NBPTS certified teachers raise student achievement on math and reading over non-certified teachers with similar levels of experience and education. We find larger effects with grade level measures for math, and comparisons between estimates from classroom measures to those aggregated at the grade level reveal that aggregation either reduces measurement error, addresses negative selection on unobservable attributes or the NBPTS effect also includes impact of positive teacher spillovers. In the following sections we address whether teacher spillovers or negative selection are the drivers of these differences.

A final note relates to generalizability of the analysis in this study. We do use a distinct sample of siblings that tend differ from the general population of students in North Carolina.[16] Nevertheless we provide additional evidence by re-estimating the preferred model given in equation (2) on the full sample of students in North Carolina public schools during this study's time period. The effects of NBPTS certification are very similar in the whole sample to sibling sample. Appendix Table 1 contains the results from the whole sample.

F.4 Are there Teacher Spillovers from NBPTS Teachers?

The larger estimates for math achievement provided by the aggregate models may reflect productivity spillovers from NBPTS teachers to their peers. NBPTS teachers have the potential to produce positive teacher spillovers, for example by exchanging lesson plans or their pedagogy

[16] Comparing means demographic characteristics by samples in table 2 we show that the sibling and twin sample is more likely to be white, less likely to be Hispanic, less likely to be limited English proficient among other smaller differences.

with other non-certified teachers in their grade, and our preferred model does not parse out this indirect effect from the direct effect of having an NBPTS teacher. To evaluate this claim, we compare whether one, two, or three or more NBPTS teachers in the grade affect the academic achievement of students whom did not have a NBPTS teacher. We do this by estimating models similar to (2) where we interchange the proportion of students in the grade with NBPTS certified teachers with dummies for whether the grade has one, two, or three or more NBPTS teachers while controlling for whether students are being taught by an NBPTS teacher. We also control for the total number of teachers in the grade to adjust for heterogeneity by grade size.

We present the results in Table VIII. For both math and reading the coefficients on the NBPTS spillover dummies are small and statistically indistinguishable from zero. These findings suggest that NBPTS teachers do not produce positive spillovers to the non-certified teachers in their grade. As a result we exclude teacher spillovers from the set of explanations.

F.5 Explaining NBPTS Certification: Signaling vs. Human Capital

To evaluate whether the effect of having an NBPTS teacher on student achievement reflects signaling or learning from the process, we estimate grade level models similar to equation (2), where in addition to the share of students in the grade taught by NBPTS teachers we also include the share of students in the grade with teachers that are not certified but will be certified in the future (proportion of pre-NBPTS certified teachers). If NBPTS certification serves solely as a signal for teacher quality, then comparing the estimate on the share of students in the grade taught by NBPTS teachers, which we will refer to as the post certification effect, to the estimate on the share of students in the grade taught by teachers who we observe get certified in future years (pre-certification effect), would show no statistically distinguishable difference in effects. On the other hand, if preparing for NBPTS certification exposes teachers to new pedagogical techniques and the preparation of the portfolio confers new skills, then the effectiveness of NBPTS certification can arise from the development of the teacher's human capital, and the post certification effect should be larger than the pre-certification effect.

We find in column one of Table IX the post-certification effect is 36% larger than the pre-certification effect for math; however, they remain statistically indistinguishable from one another. On the contrary for reading, the effect is 46% larger pre-certification and statistically insignificant from the post certification effect. Overall the evidence appears to support the signaling hypothesis.

F.6 The Effect of NBPTS Teachers by Elementary and Middle School Grades

Several factors such as teacher professional development and the degree of difficulty of required teaching content raises the possibility that the contribution of NBPTS certification may vary by specific grade. To determine whether such heterogeneity exists, we estimate equation (2) with the right hand side fully interacted with a middle school indicator.[17] Column one of Table X shows that the effect of NBPTS certification on elementary math is 0.013 of a standard deviation and statistically indistinguishable from zero. On the contrary for middle school (column three), we find that NBPTS teachers are 0.057 of a standard deviation more effective than non-certified teachers. In columns two and four we provide results for reading achievement. The elementary sample once again produces a small and statistically insignificant effect. For middle school, on the other hand we do find that NBPTS teachers significantly improve reading outcomes above non-certified teachers by 0.02 of standard deviation.

F.7 Does School Poverty Influence the Effectiveness of NBPTS Teachers?

Lastly we evaluate whether the value of an NBPTS teacher varies by the level of poverty experienced at different schools. We hypothesize that at lower SES schools home resources are

[17] Grade 4-5th are considered elementary school, and grades 6th -8thare considered middle school.

provided at lower rates and therefore NBPTS teachers may have a larger potential to improve outcomes, particularly in reading. We use free and reduced lunch status to classify if the schools is a low socioeconomic status (SES) school or not, and we then run equation (2) fully interacted with the low SES school indicator. Low SES schools are characterized as schools with 75% or more of their students on free or reduced lunch and non-low SES schools are schools with less than 75% of their students on free or reduced lunch. The results are robust to alternative classification thresholds.

Results are presented in Table XI. Columns one and three show that on average a student in both a low SES or non-low SES school achieves test score gains of 0.043 of a standard deviation in math from being in a grade with all NBPTS certified teachers. For reading we observe a small positive significant effects of 0.015 of a standard deviation for non-poor schools. At poor schools, however, we observe a negative and statistically insignificant effect of NBPTS certification. However the standard error for the latter is considerably large such that we cannot reject this effect from the estimate for the non-poor schools. The main findings from Table XI suggest that there does not appear to be evidence of heterogeneity of effectiveness of NPBTS teacher by the poverty composition of schools.

G. Conclusion

The National Board for Professional Teaching Standards certification as a voluntary credential offers several potential pathways for linkages to teacher quality. Notable work on teacher quality illustrates that within schools teachers are one of the most important factors linked to student outcomes, and identifying superior teachers is an important priority for schools and districts. In this study we credibly identify the effect of an NBPTS certified teacher exploiting several sources of variation including within twins, within siblings at different schools, and across cohorts within schools. The analysis in this study demonstrates that NBPTS teachers are indeed more effective at improving student academic achievement on both math and reading assessments.

We find that NBPTS teacher on average raise student achievement by 0.04 of a standard deviation in math and 0.013 of a standard deviation in reading when aggregating to the grade and making comparisons across cohorts within a school-by-year. Compared to classroom level estimates derived from within family variation and within school variation, our preferred estimates are similar for reading and about 15- 30% larger in math. Moreover, including sibling-year fixed effects in addition to school-by-year fixed effects does not substantially affect estimates derived at the classroom level, which suggests that either included controls capture the dimensions on which students are being sorted within schools, there is potential for selection within family, or within family peer effects suppress the true effect.

Several reasons could explain why aggregation produces larger results than classroom level models including measurement error, teacher spillovers, and sorting on unobservable characteristics. Attempts at assessing teacher spillovers as an explanation reveal that they are not a driving factor, as we find no evidence that NBPTS teachers improve the effectiveness of non-NBPTS teachers. Other potential explanation include student unobserved heterogeneity or reduction in measurement error. The differences are small that even if the estimates from classroom models are biased, the bias is negligible.

Our analysis additionally reveals considerable heterogeneity in effectiveness by middle school and elementary school. While elementary NBPTS certified teachers only marginally improve their students' test-scores, certified middle school teachers show large improvements with the most substantial coming from middle school math teachers. In addition to heterogeneity by type of schooling, we also investigate heterogeneity by school poverty. We find no evidence that NBPTS teachers are more effective at schools with a large proportion of students on free-reduced lunch.

Although the black box of how teachers raise achievement may remain murky, the overall evidence supports the claim that NBPTS certification can be explained by signaling as opposing to human capital. We show that good teachers separate themselves on their measures of teaching ability, and pursuing NBPTS certification does not improve their human capital. Teacher characteristics provide auxiliary evidence in support of the claim that NBPTS certification is a signal of teacher quality, and we find for example that NBPTS certified teacher tend to have higher PRAXIS scores and are more likely to have advanced degrees than non-NBPTS teachers.

From a policy perspective, it is unclear if NBPTS certification is a cost effective approach to raising achievement. Notable work on the relationship between classroom size and academic achievement finds smaller classes do raise achievement-0.020 of standard deviations per student (Krueger and Whitmore, 2001; Krueger, 2003; Hanushek, 1999; Hanushek, 2002). Yet, reducing class size is often not implemented because of staffing costs. To assess whether teachers with NBPTS certification are an economical way of raising the quality of instruction we crudely quantify whether the benefits to students as measured by the present value of future earnings gains offset the certification salary premium. In North Carolina a certified teacher with a Master's degree and 14 years of experience (the average teacher experience in our sample) received an additional $5,240 in wages in fiscal year 2011-2012 (North Carolina Department of Instruction). To calculate the present value of future earnings gains to students we use the earning returns from a one standard deviation increase in teacher value-added estimated in

Chetty et al.[b] (2014). Using our aggregate estimate of having a certified teacher in math reveals on average an increase in earnings by age 28 of $117 or 0.33%. Assuming the percentage impact remains constant over the lifecycle and a 3% discount rate, the present value of future earnings gains at 12 years of age, the average age in our sample, then aggregated to the class equates to about $48,000:[18]

$$(23 \text{ students}) * \sum_{t=16}^{56} \frac{\$117}{(1.03)^t} = \$48{,}000$$

The value of NPBTS teachers is substantial, and importantly offsets the certification wage premium. Policies that make use of NBPTS certification whether to identify or retain good teachers, are an economical way of raising the quality of instruction that may potentially provide large long run economic and social benefits.

[18] We use the same discount rate as Chetty et al.[b] 2014.

Table I
Descriptive Statistics of National Board Teachers vs. Non National Board Teachers[a]

Teacher characteristics	NB Certified Mean	Not Certified Mean	P-Value
Teacher experience	14.06	11.22	0.00
Proportion with advanced degree	0.51	0.30	0.00
Proportion white	0.92	0.82	0.00
Proportion black	0.06	0.16	0.00
Proportion Asian	0.00	0.01	0.00
Proportion female	0.94	0.88	0.00
Standardized PRAXIS	0.18	-0.09	0.00
Observations	35032	199349	
School characteristics			
Proportion free or reduced lunch	0.47	0.53	0.00
Proportion Black	0.23	0.30	0.00
Proportion White	0.59	0.50	0.00
Proportion Hispanic	0.12	0.13	0.00
Observations	35032	199349	

Notes: [a]Source Data: NCERDC

Table II
Career Attributes of National Board Certified Teachers[b]

Teacher Attributes Year Certified	Mean		
Teacher experience	11.36		
Proportion with Bachelors	0.55		
Proportion with advanced degree	0.45		
Observations	6060		
School Attributes Before and After NBPTS Certification			
	Before Mean	**After** Mean	P-Value
Proportion free or reduced lunch	0.45	0.46	0
Proportion Black	0.25	0.22	0
Proportion White	0.59	0.61	0
Proportion Hispanic	0.11	0.11	0
Proportion Other	0.06	0.06	
Observations	4888	18247	

Notes: [a]Source Data: NCERDC

Table III
Descriptive Statistics for Full Student, Sibling, and Twin Samples[a]

	Full Sample	Sibling Sample	Twin Sample
Classroom Characteristics	Mean	Mean	Mean
Class size	23.405	23.436	23.726
Teacher experience	11.799	11.983	11.934
Teacher has Masters plus	0.301	0.303	0.309
National Board certified	0.121	0.124	0.128
Student Characteristics			
Math	0.032	0.057	0.124
Reading	0.023	0.012	0.099
Economically disadvantaged	0.501	0.524	0.449
Black	0.262	0.269	0.250
White	0.562	0.612	0.656
Hispanic	0.103	0.054	0.041
Female	0.497	0.499	0.514
Disabled	0.108	0.110	0.125
Limited English Proficient	0.049	0.025	0.019
Age	12.232	12.231	12.187
Observations	2874050	628963	41174

Notes: [a]Source Data: NCERDC

Table IV

The Effect of Having an NBPTS Certified Teacher on Math and Reading Achievement for the Sibling Sample[a,b]

	(1)	(2)	(3)	(4)	(5)
Panel A. Math Achievement					
Student has NBPTS teacher	0.047***	0.036***	0.032***	0.033***	0.027***
	(0.003)	(0.003)	(0.005)	(0.003)	(0.005)
Observations=628963					
Panel B. Reading Achievement					
Student has NBPTS teacher	0.017***	0.013***	0.015***	0.013***	0.019***
	(0.002)	(0.002)	(0.004)	(0.002)	(0.005)
Observations=617683					
School-by-Year FE	No	Yes	Yes	No	No
School-by-Grade-by-Year FE	No	No	No	Yes	Yes
Sibling-by-Year FE	No	No	Yes	No	Yes

Notes: [a]Each cell estimate is derived from a separate model for the full sibling sample. Models include student, teacher and classroom controls. Student controls include a cubic in lagged test scores in both math and reading, age, birth order, spacing of siblings, and indicators for race, gender, disability, limited English proficiency indicator, and economically disadvantaged indicator. Teacher controls include experience dummies and an indicator for advanced degree. Peer controls include mean subject specific lagged test-score, class-size, mean age, proportion non-white, proportion limited English proficient, proportion disabled and proportion economically disadvantaged. Models in columns 1-3 and 5 also include grade-by-year fixed effects. Standard errors found in parentheses are clustered at teacher-by-year level (* p<0.05, ** p<0.01, *** p<0.001).
[b]Source Data: NCERDC

Table V
The Effect of Having an NBPTS Certified Teacher on Math and Reading Achievement for the Twin Sample[a,b]

	(1)	(2)	(3)
Panel A. Math			
Student has NBPTS teacher	0.042***	0.0152	0.029**
	(0.009)	(0.011)	(0.011)
Observations= 41176			
Panel B. Reading			
Student has NBPTS teacher	0.025*	0.0273*	0.025*
	(0.008)	(0.011)	(0.011)
Observations= 40827			
School-by-Grade-by-Year FE	No	Yes	No
Twin-by-Year FE	No	No	Yes

Notes: [a]Each cell estimate is derived from a separate model. Models include student, teacher and classroom controls. Student controls in the first two columns include a cubic in lagged test scores in both math and reading, age, birth order, spacing of siblings, and indicators for race, gender, disability, limited English proficiency indicator, and economically disadvantaged indicator. In the last column we only included controls that vary within in siblings. Teacher controls include experience dummies and an indicator for advanced degree. Peer controls include mean subject specific lagged test-score, class-size, mean age, proportion non-white, proportion limited English proficient, proportion disabled and proportion economically disadvantaged. Standard errors found in parentheses are clustered at teacher-by-year level (* p<0.05, ** p<0.01, *** p<0.001).
[b]Source Data: NCERDC

Table VI
Predicting the Probability of Exposure to NBPTS Teachers with Student Characteristics[a,b]

	(1)	(2)	(3)	(4)
Economically disadvantaged	-0.011***	-	-	-0.032
	(0.002)	-	-	(0.057)
Female	0.001	0.004*	-0.005	0.021
	(0.001)	(0.002)	(0.006)	(0.057)
Behavioral disability	-0.029*	-0.042*	-0.005	-0.039
	(0.011)	(0.015)	(0.057)	(0.313)
Other disability	0.004	-0.000	-0.020*	-0.024
	(0.002)	(0.003)	(0.009)	(0.068)
Limited English proficient	0.007	-	-	-0.077
	(0.005)	-	-	(0.106)
Black	0.001	-	-	-0.182*
	(0.002)	-	-	(0.082)
Hispanic	0.002	-	-	-0.032
	(0.004)	-	-	(0.114)
Other	0.003	-	-	-0.096
	(0.003)	-	-	(0.120)
Lag math	0.017***	0.016***	0.023***	0.024
	(0.001)	(0.002)	(0.004)	(0.024)
Lag reading	0.005***	0.003	0.006	-0.012
	(0.001)	(0.002)	(0.004)	(0.031)
School-by-Year FE	Yes	No	No	Yes
Siblings-by-Year FE	No	Yes	No	No
Twins-by-Year FE	No	No	Yes	No
Observations	288395	290479	20288	288395

Notes: [a]Column 1-3 presents results from a linear probability model predicting the probability of taking math with an NPBTS teacher. Column 4 predicts the proportion of students in the grade exposed to NBPTS teachers in math with grade aggregates of student characteristics. Additional controls in 1, 2 and 4 include age and grade-by-year dummies. The sample is constrained to school years that contain both NBPTS certified and non-certified teachers. Cells with dashes imply that the particular student characteristic was not included in the model, because it did not vary. Standard errors provided in parentheses are clustered at the teacher-by-year level in columns 1-3 and school-by-grade-by-year level in column 4 (* p<0.05, ** p<0.01, *** p<0.001).
[b]Source Data: NCERDC

Table VII
The Effect of NBPTS Certification at Grade Level on Math and Reading Achievement for the Sibling Sample[a,b]

	(1)	(2)	(3)
Panel A: Math Achievement			
Proportion of students in grade taught by NBPTS teachers	0.065***	0.041***	0.042***
	(0.006)	(0.006)	(0.009)
Observations=628963			
Panel B: Reading Achievement			
Student has NBPTS teacher	0.020***	0.012**	0.009
	(0.004)	(0.004)	(0.008)
Observations=617683			
School-by-Year FE	No	Yes	Yes
Sibling-by-Year FE	No	No	Yes

Notes: [a]Each cell estimate is derived from a separate model. Student controls in models include a cubic expansion in lagged test scores in both math and reading, age, birth order, spacing of siblings, race indicators, gender indicator, disability indicator, limited English proficiency indicator, and economically disadvantaged indicator. Teacher controls include proportion of students being taught by buckets of different experienced teachers and proportion of students being taught by teachers with an advanced degree. Grade peer controls include mean subject specific lagged test-score, mean class-size, mean age, proportion non-white, proportion limited English proficient, proportion disabled and proportion economically disadvantaged. All models also include grade-by-year fixed effects. Standard errors found in parentheses are clustered at the school-by-year-by-grade level (* p<0.05, ** p<0.01, *** p<0.001).
[b]Source Data: NCERDC

Table VIII
Assessing NBPTS Teacher Spillovers[a,b]

	Math (1)	Reading (2)
One NBPTS teacher in grade	-0.002	-0.002
	(0.003)	(0.002)
Two NBPTS teachers in grade	-0.000	-0.002
	(0.005)	(0.004)
Three plus NBPTS teachers in grade	-0.011	-0.001
	(0.009)	(0.006)
School-by-Year FE	Yes	Yes
Observations	628963	617683

Notes: [a]Each column presents results from the same model. All models include the student, teacher, and grade controls, which were included in the models specified in Table VII. In addition all models include the total number of teachers in the grade and a dummy for whether the student's subject specific teacher is NBPTS certified. Standard errors clustered at the school-by-grade-by-year are presented in parentheses (* $p<0.05$, ** $p<0.01$, *** $p<0.001$).
[b]Source Data: NCERDC

Table IX

Does NBPTS Certification Improve Teacher Productivity[a,b]?

	Math (1)	Reading (2)
Proportion pre NBPTS teachers in grade	0.033*	0.027**
	(0.012)	(0.009)
Proportion NBPTS teachers in grade	0.042***	0.013**
	(0.006)	(0.004)
School-by-Year FE	Yes	Yes
Observations	628963	617683

Notes: [a]Each column presents estimates from the separate model. All models include the student, teacher, and grade controls, which were specified in Table VII. Standard errors clustered at the school-by-grade-by-year are presented in parentheses (* p<0.05, ** p<0.01, *** p<0.001).

[b]Source Data: NCERDC

Table X
The Effect of NBPTS Certification on Math and Reading Achievement Stratified by Elementary and Middle School[a,b]

	Elementary School		Middle School	
	Math	Reading	Math	Reading
	(1)	(2)	(3)	(4)
Proportion NBPTS teachers in grade	0.013	-0.003	0.057***	0.019*
	(0.009)	(0.007)	(0.012)	(0.009)
Observations	628963	617683	628963	617683
School-Year FE	Yes	Yes	Yes	Yes

Notes: [a]Cell results in columns 1 and 3 and in 2 and 4 are estimated in the same model. All models include the student, teacher, and grade controls, which were specified in Table VII. The elementary school sample contains students in 4th and 5th grades, and the middle school sample contains students in 6th, 7th, and 8th grades. Standard errors clustered at the school-by-grade-by-year are presented in parentheses (* p<0.05, ** p<0.01, *** p<0.001).
[b]Source Data: NCERDC

Table XI

The Effect of NBPTS Certification on Math and Reading Achievement stratified by School SES[a,b]

	Low SES		Non-Low SES	
	Math	Reading	Math	Reading
	(1)	(2)	(3)	(4)
Proportion NBPTS teachers in grade	0.044*	-0.008	0.042***	0.014**
	(0.022)	(0.018)	(0.006)	(0.005)
Observations	586775	575402	586775	575402
School-by-Year FE	Yes	Yes	Yes	Yes

Notes: [a]Cell estimates in columns 1 and 3, and 2 and 4 are estimated in the same model. All models include the student, teacher, and grade controls, which were included in the models specified in Table VII. Low SES schools include schools with 75% or more of their students and free and reduced lunch. Standard errors clustered at the school-by-grade-year are presented in parentheses (* p<0.05, ** p<0.01, *** p<0.001).

[b]Source Data: NCERDC

III. THE PERSISTENT EFFECTS OF NATIONAL BOARD CERTIFIED TEACHERS

A. Introduction

Many studies including Chapter II show that students of NBPTS teachers outperform students of non-certified teachers. This body of work concludes that the National Boards identify effective teaching (Clotfelter et al, 2007; Cowan and Goldhaber, 2015; Harris and Sass, 2009). Few studies, however, demonstrate whether the effects of NBPTS teachers persist overtime (Jacob et al. 2010). And while there has been a plethora of evidence that teacher effects matter even two years (Chetty et al. 2014[b]; McCaffey et al. 2004; Rothstein, 2010; Kane and Staiger 2008), there has been almost no work addressing whether certified teachers have lasting impacts. The literature has shown that the certification is a good proxy of teacher quality when testing with contemporaneous student outcomes. However, a good measure of teacher effectiveness should have implications for long-term student outcomes. Since education attainment is a cumulative process, it requires that there is persistence of knowledge overtime. Of course not all skills learned today are needed for tomorrow's tasks, but some amount are, and therefore effective teaching should drive permanent knowledge gains and not just transient ones. As such, it is important address whether the effects of NBPTS teachers persist into the future.

To test for persistence of NBPTS teacher effects on future student outcomes I continue to use the student-teacher matched longitudinal administrative data from North Carolina. The main outcomes of interest are standardized math and reading test-scores one year out. Treatment will be subject specific and measured two different ways. First, I measure whether the student in time t has an NBPTS teacher with an indicator variable, and I run models accounting for a rich set of controls including test-scores from the previous year and school-by-grade-by-year fixed effects in year of treatment thus accounting for non-random selection to schools, unobserved differences among cohorts, and unobserved differences in the distribution of school input between grades.

While these models run with contemporaneous outcomes have shown to have small biases due to non-random sorting of students and teachers within schools (Chetty et al., 2014[a] ;Chapter II), to directly address classroom selection bias I also run models where I aggregate the treatment variable to the school-by-grade-by-year level and account for a rich set of controls. To account for fixed differences between grades within schools and selection to schools I include school-by-grade fixed effects. This is particularly important to adjust for when using cohort variation as these models are susceptible to contamination of control cohorts in the future period when the outcome is realized. Thus adjusting for fixed differences between grades will make cleaner comparisons, however as long as inputs are dynamically adjusted within schools, there are still possibilities that the control group is treated. Classroom level models with school-by-grade-by-years fixed effects do not suffer from this shortcoming.

My findings show that certified teachers have lasting impacts one year out by 0.008-.015 of a standard deviation in math and 0.007-0.012 of a standard deviation in reading using across classroom variation. Relative to contemporaneous effects, these effects represent full persistence for reading, and 30% persistence for math. The different findings among math and reading have three likely determinants. First, mathematics curriculum may have less overlap across grades relative to reading. Second, even if the overlap is identical, reading is complementarily reinforced outside of the classroom. Last, certified teachers teaching math may be more likely to teach transient skills or to the test relative to certified instructors teaching reading.

Grade level models on the other hand show that neither math nor reading certified teachers' effects persist. There are two potential explanations. First, classroom models may suffer from positive selection. Alternatively, there may be an inverse relationship between the qualities of inputs from year to year due to capacity constraints of schools that lead to the control cohorts to get treated in the year the outcome is realized. This relationship would downward bias our estimates. Preliminary evidence shows that the share of students exposed to certified teachers in the contemporaneous period is inversely related to the share of certified teachers in following lead year within a school, which supports the latter hypothesis.

Since studies show the effectiveness of the certified teachers varies by grade level, I also test for heterogeneity of the effect by middle (6^{th} and 7^{th}) and elementary (4^{th} and 5^{th}) grades (Chapter II; Cowan and Goldhaber 2015). Classroom level models show that the persistence of certified teachers is similar among elementary and middle school in terms of depreciation rates. However, just as other studies have found, certification in middle school tends to do a better job at uncovering effective teachers. Grade level models find inconclusive results for persistence although the contemporaneous models show a similar pattern to the classroom level models.

B. Empirical Strategy

Estimating the persistent effects of NBPTS teachers on student achievement in future time periods is a complex feat. As in the contemporaneous model world we face several potential concerns for bias. The first and foremost is that of student and teacher sorting on unobservables to schools and classrooms both within and across time. In addition, we also face problems with contemporaneous shocks that have permanent effects on learning and unobserved differences across cohorts within schools.

To overcome potential selection concerns I use several empirical strategies to isolate as good as random variation in the students' exposure to NBPTS certified teachers. The first strategy uses across classroom variation in the exposure of certified teachers accounting for a rich set of controls including lagged measures of academic achievement and school-by-year-grade fixed effects. The empirical model is shown in equation (1).

$$(1)\ A_{ig+1t+1} = f\left(A_{ig-1t-1}\right)\lambda + \beta NBPTS_t + X_{it}\,\delta + \bar{X}_c\pi + T_t\rho + \gamma_{gt} + \theta_{sgt}\ + \epsilon_{ig+1t+1}$$

The outcome of interest $A_{ig+1st+1}$ is the subject specific student achievement score standardized by grade and year for student *i* in grade *g+1* and time *t+1*. We model achievement in *t+1* as a function of a cubic expansion in prior test scores in both math and reading taken in *t-1,* $f\left(A_{ig-1t-1}\right)$; an indicator variable for whether the student had an NBPTS teacher for the particular subject in time *t*; a vector of student characteristics, X_{it} , measured at time *t* that include age, indicators for race, gender, disability status, limited English proficiency indicator, and economically disadvantaged; a vector of classroom characteristics, $\bar{X}_c$, measured at time *t* that include mean subject-specific lagged test-score and age, class-size, and proportion non-white, limited English proficient, disabled and economically disadvantaged; a vector of teacher characteristics of students i subject specific teacher at time *t*, T_t, that includes experience dummies and an indicator for advanced degree. I also include grade-by-year dummies to account for changes in curriculum and tests,γ_{gt}, and school-by-grade-year fixed effects θ_{sgt}.

These models address the most pressing concerns including sorting to schools on unobservables, school level shocks, time-varying and fixed differences among grades within schools, and sorting on observables within schools. In Chapter II I show that the relative bias for the remaining sorting on unobervables within schools were small for math and nearly non-existent for reading. In addition, I run a number of specification checks including exchanging school-by-grade-by-year fixed effects with school-by-grade fixed effects to just account for unobservable fixed differences between grades within schools, including both school-by-grade and school-by-year fixed effects, and exchanging them both for teacher-by-year fixed effects in *t+1* to account for unobservable sorting to classrooms. All specifications are run separately for math and reading achievement, and all standard errors are clustered to the teacher-by-year level in time *t*.

While these set of models have shown to do a good job at accounting for unobserved differences between classrooms, I can directly account for these differences with a second method which employs across cohort variation in the exposure to certified teachers with-in a school and grade in time *t*. The empirical specification is shown in equation (2) where the estimate of interest, β, is on the proportion of students in the grade taught by NBPTS teachers in time *t* and school *s*, $\overline{NBPTS}_{gst}$.

$$(2)\ A_{ig+1t+1} = f(A_{ig-1t-1})\lambda + \beta\overline{NBPTS}_{gst} + X_{it}\,\delta + \bar{X}_{gst}\pi + \bar{T}_{gst}\rho + \gamma_{gt} + \theta_{sg}\ + \epsilon_{ig+1t+1}$$

Similar to the classroom level models I include lagged achievement using cubic expansions in prior test scores in both math and reading, $f(A_{ig-1t-1})$, the exact same student controls, X_{it} , grade-by-year fixed effects, γ_{gt}, and school-by-year fixed effects, θ_{st}. The models differ, however, as the variable of interest, $\overline{NBPTS}_{gst}$, other teacher credentials, $\bar{T}_{gst}$, are aggregated to the school-by-grade-by-year level at time *t* such that each student-by-year observation receives the grade level mean of the variable in question. Instead of including classroom characteristics, we include grade characteristics of the same variables, $\bar{X}_{gst}$. In addition, I run a number of specification checks including exchanging school-by-grade fixed effects with school-by-year fixed effects to account for school shocks with permanent impacts on learning, and including both of these fixed effects. Further all specifications are run separately for math and reading achievement, and standard errors are clustered to the school-by-grade-by-year level in time *t*.

In this model the contaminated classroom variation is subsumed in the error term and it is mechanically uncorrelated with the cohort variation, thus eliminating one source of bias. The addition of the school-by-grade fixed effects account for sorting to schools and fixed differences between cohorts and input quality across grades. To the extent that these differences vary overtime these estimates could still be contaminated with bias even if these differences are random. I explore for this potential threat by running models similar to (1) but where I exchange the achievement outcome with whether the student the student has an NBPTS teacher in *t+1*.

C. Data

This study uses school administrative data with matched teachers to student records from North Carolina public schools housed at the North Carolina Education Research Data Center (NCERDC) for third through eighth graders from 2006 to 2013. Student variables include race, gender, economically disadvantaged status, limited English proficiency status, disability status, age and end-of-grade standardized test-scores in both math and reading. Teacher characteristics are pulled from teacher pay records, and include years of experience, educational attainment, and national board certification status. Since I use lagged achievement models, and standardized testing does not begin until students reach third grade in North Carolina, I use third grade achievement as the baseline measure for lagged achievement for the students in 4^{th} grade. In addition, our estimation sample begins with the cohort from 2007 as we use lagged achievement from 2006 as the baseline achievement for 2007.

Finally the analysis sample consists of students that spend at least three consecutive years in a public school in North Carolina between 2006-2013. This is because I will be running models that predict subject specific test scores in the year following being taught by a certified teacher, but I will also account for past inputs with lagged achievement variables from the year prior to having the teacher. This also means that 8th graders will not show up in my analysis sample as that is the last year I observe students with standardized tests.

To match students to their subject specific teacher and peers, I use course-membership files and group students on year, school, course title, semester and section. This procedure identifies the students' subject specific classroom. I restrict the analysis to math and reading classroom(s) and run models separately by subject.

In Table XII I show descriptive statistics for the full sample of 4th-7th grade students in the study time period and the analysis sample that makes the additional restriction of students with at least three consecutive years of data.[19] Relative to the full sample, an average student in the analysis sample has higher academic achievement in both reading and math by about .03 of a standard deviation respectively. On all other student and classroom characteristics the samples are balanced.

The small differences between the full and analysis samples potentially signals that the analysis sample suffers from non-random attrition. However a simple comparison of the attrition rates between students with certified and non-certified teachers reveals that the rates are identical: 14.93% compared to 15.02% respectively with their differences being statistically insignificant. In addition to testing for whether attrition rates differ by treated and control groups, I test whether treated and control groups who attrite differ in observable demographics after adjusting for selection to treatment. To test for these differences I predict demographic characteristics with a fully interacted model of whether the student leaves the sample with an

[19] I only look at 4th – 7th for the full sample to make comparisons across similar groups of students. The analysis sample will never have 8th graders since we do not see student in 9th grade with a standardized test score.

indicator for whether the student has an NBPTS teacher and the selection controls containing cubic polynomials in lagged test scores of both subjects and school-by-year fixed effects. Results can be found in Table XIII. All but two of the adjusted mean differences are statistically insignificant, and of the two that were significant the difference was small. Specifically students who leave the sample and were treated with a certified teachers are less likely to be economically disadvantaged by 0.7 of a percentage and more likely to be in the other race category by 0.3 of a percentage point. Taken together, the two tests provide evidence that attrition bias is an unlikely source of bias in the estimation of the persistence effects of NBPTS teachers.

D. Results

D.1 Does the Effect of NBPTS Teachers Persist One-Year Out?

D.1.1 Estimates Based on Classroom Variation

Chapter II provides evidence that students of NBPTS teachers outperform students of non-certified teachers in the year they have these teachers. In this chapter, I am interested in understanding whether these effects persist one year out. Table XIV provides results for math and reading achievement one year out (Panel A.) and contemporaneously (Panel B.) from running models with the framework described by equation (1). Columns 1-5 provide estimates for math and 6-10 for reading.

Students of certified students continue to outperform students of non-certified students one year out by 0.009 of a standard deviation in specification accounting for school-by-year fixed effects at the time of treatment found in Panel A. column one. Since there is potential for non-uniform resource allocation across grades, column two exchanges school-by-year fixed effects with school-by-grade fixed effects thereby accounting for fixed differences between grades within schools. The estimate continues to show that the effects of certified teachers persists with students of certified continue to outperform those of non-certified by 0.011 of a standard deviation. The third column includes both school-by-year and school-by-grade fixed effects, and the estimate is slightly larger to 0.013 of a standard deviation. Column four accounts explicitly for differences in cohorts and resource allocation across grades within school by including school-by-year-by-grade fixed effects and the estimate increases to 0.015 of standard deviation. While this estimate is the largest it is also statistically indistinguishable from those in columns two and three. Further, while I control for many potential differences with school-by-grade-by-year fixed effects, the remaining variation in column four is driven by variation across classrooms within in grades, and to the extent that there remains positive selection on unobservables, this estimate could be biased up. Therefore in column five I exchange the school-by-grade-by-year fixed effects with teacher-by-year fixed effects of the teacher and year student has in *t+1*. This model more directly accounts for classroom selection by comparing students who end up with teachers with the same teacher quality in *t+1* regardless of treatment status. Here I find that the estimate reduces to 0.008 of a standard deviation.[20]

Similar patterns are shown for reading in columns 6-10. Column six, which runs equation (1) with school-by-year fixed effects, shows that students taking reading with NBPTS certified teachers continue to outperform students of non-certified teachers by 0.009 of a standard deviation. When I exchange school-by-year fixed effects with school-by-grade fixed effects the estimate increases to 0.011 of standard deviation, and this result is robust to including both fixed effects and school-by-grade-by-year fixed effects. The last column uses teacher-by-year fixed

[20] Teacher quality in t+1 is a potential outcome of treatment with a certified teacher in the prior year and could be a bad control. Students who get matched to the same quality teacher in t+1 could differ in unobservable ways if treatment impacts the probability of getting a higher quality teacher. Students who were treated in the prior year are likely negatively selected since untreated students would have to be more special to end up with the same quality teacher.

effects of the teacher and year when the outcome is realized and the estimate reduces to 0.007 of a standard deviation.

Both math and reading results show that certified teachers have lasting impacts one year out. While the magnitudes differ slightly based on the specification, the results are qualitatively the same. To understand how much teacher effects persist I compare contemporaneous effects in Panel B. to the persistent effects of Panel A. It can be seen that the NBPTS effects for math one year out depreciate by about 65%, and this holds across all specifications. For example, column one of Panel B. shows the effect of having an NBPTS teacher impacts math achievement on average by 0.029 of a standard deviation. The comparable model for the long-run effect however shows that NBPTS effects only persist by 0.009 of standard deviation, which is about 65% smaller than the contemporaneous effect.

Comparing contemporaneous NBPTS effects for reading to estimates for persistence presents a slightly different story. Within specification but across outcomes we see that the estimates are nearly identical. Or the effects of NBPTS reading teachers do not depreciate one year out. The non-depreciation suggests that effective teachers more easily improve permanent reading skills rather than math skills. However, it is also important to consider that mathematics curriculum has likely less overlap across grades than reading curriculum.

D.1.2 Estimates Based on Cohort Variation

The alternative estimation strategy to identify the persistence of NBPTS teacher effects relies on isolating across cohort variation as described by equation (2). Table XV, Panel A. holds results from this strategy for math (columns 1-3) and reading (columns 4-6). Column one starts with a model that accounts for school-by-year fixed effects for the school and year at time of treatment. With this model the persistence effects for math in $t+1$ is statistically insignificant

-0.007 of a standard deviation. While this estimate suggests that there is no persistent effect of certified teachers for math achievement, this model uses variation that is susceptible to treatment of control cohorts in the year the outcome is realized. Since there are few certified teachers in schools to begin with they likely are in one or two grades, and thus in grade *g+1* of the year the outcome is realized, the control group will get treated leading to a downward bias. To partially adjust for this issue I run equation (2) explicitly, which instead controls for school-by-grade fixed effects rather than school-by-year fixed effects. Results are presented in column two. The estimate of persistence increases to 0.00 of a standard deviation, however the result is imprecisely estimated such that one cannot reject any of the estimates of the classroom variation strategy shown in Table XIV. Column three includes both school-by-year and school-by-grade fixed effects and the result remains qualitatively the same.

Reading results show a similar pattern. Students of NBPTS teachers have lower test scores on average in the following year by an insignificant 0.004 of standard deviation as estimated by model with school-by-year fixed effects in column four. Exchanging school-by-year fixed effects with school-by-grade fixed effects flips the estimate to a positive and insignificant effect of 0.005 of standard deviation in column 5five The last column includes both fixed effects and the estimate reduces to zero however imprecisely estimated.

Panel B. of Table XV presents results for contemporaneous models, where math and reading outcomes are now measured in the same year the student is treated. The right hand side of the models remains identical to the models testing for persistence. Contemporaneous models, as expected, show that certified teachers are more effective teachers in both math and reading subjects. While the estimates tend to be slightly smaller relative to estimates presented in chapter

II, they are qualitatively the same.[21] Depending on the specification the contemporaneous models find that NBPTS teachers are more effective than non-certified teacher by 0.022-0.048 of a standard deviation for math and 0.004-0.011 for reading. These estimates align with contemporaneous effects derived from estimation models relying on across classroom variation in Table XIV.

D.2 What Drives Differences Between the Persistent Effects of NBPTS Teachers of Classroom and Cohort Models

Comparisons of persistent effects of NBPTS teachers between Tables XIV and XV reveal slightly different conclusions. Models relying on across classroom variation show that NBPTS teacher effects persist one year out for both math and reading. On the contrary, estimates derived from cohort models in Table XV show the persistence of NBPTS teacher effects are small, sometimes negative, and wildly imprecise. There are two potential explanations for the observed differences. First, classroom models may potential suffer from positive selection. Alternatively there is an inverse relationship between the qualities of inputs from year to year due to capacity constraints of schools that would lead to contaminated control cohorts.

The first potential explanation is unlikely as classroom models accounting for teacher quality in year *t+1* still produce positive effects for both math and reading. Since these models compare treated and non-treated students who are in the same classroom in year *t+1*, it accounts explicitly for selection to classrooms. Further, because treatment at time *t* affects the quality of the student's following math or reading teacher, the estimate from this model is still subject to selection bias, however, the bias is likely to be negative. Untreated students at time *t* would have

[21] The analysis sample used here does not include 8th graders. However, the sample in chapter 2 did. Since NBPTS teachers were found to be more effective in middle school grades, dropping the 8th grade student sample could explain why the estimates tend to be smaller.

had to be especially strong academically to end up with the same quality teacher as treated students in *t+1*. As such these estimates are lower bounds.

To test whether control cohorts are contaminated with treatment in year *t+1* I predict the probability of having an NBPTS teacher at time *t+1* with an indicator for whether the student has an NBPTS teacher at time *t*. The model also includes all controls specified in equation (1) and school-by-year fixed effects. I run this model on the full analysis math sample and on a constrained sample of students that attend schools where in year *t* each grade in the school has at least one NBPTS teacher that teaches math. If control cohorts are treated in year *t+1* then the results from the full sample would show that students treated with NBPTS teachers in year t are on average less likely to have an NBPTS teacher in the following year relative to non-treated students. In addition, the constrained sample estimate should be close to zero or even positive, because in this sample every student has an opportunity to have a certified teacher. The negative coefficient of the full sample is than a result of students in the control group who did not have a chance to have a certified teacher in year *t*, but do in *t+1*. Results to these models are found in Table XVI where column one holds the full sample estimate, and column two the constrained sample estimate. In the full sample students treated with certified teachers in *t* are 5.4 percentage points less likely to have a certified teacher for math the following year. Off a mean of 12%, this estimate is equivalent to a 45% reduction in the likelihood of having a certified teacher the following year. On the contrary students treated with certified teachers in *t* and who have a chance to have a certified teacher each year are no more likely to have a certified teacher the following year. The estimates follow the pattern indicating that control cohorts are contaminated with treatment.

The last two columns of Table XVI provide evidence that cohort models that use variation overtime rather than the cross-section are preferred. I predict the share of students exposed to NBPTS teachers when students are in grade $g+1$ with the share of students exposed to NBPTS teachers when in grade g and vary between models with school-by-year and school-by-grade fixed effects.[22] Since school-by-grade fixed effect models account for fixed differences between grades within schools, I expect that these models reduce the contamination bias most. Thus, the relationship between the shares of students exposed to NBPTS teachers in t and $t+1$ estimated with school-by-grade fixed effects should be in absolute value smaller than of the estimate from school-by-year fixed model. The school-by-year fixed effect model in column three predicts that being in a grade with full exposure to NBPTS teachers reduces the likelihood that one is exposed to an NBPTS teacher in the following grade by 32 percentage points. On the contrary, the estimate in column four, which uses variation across time in the share of students exposed to NBPTS teachers in the grade, is marginally positive but statistically insignificant.[23]

Taken together these results provide evidence that school-by-year fixed effects suffer greatly from treatment contamination of control cohorts, while school-by-grade fixed effects models seem to alleviate much of the contamination with a caveat that these models tend to be imprecise due to the large reduction in identifying variation. Importantly, this bias does not only impact estimates employing cohort variation, but also some of the results estimated by employing classroom variation although to a lesser extent. Estimates that do not suffer from this bias consist of those that are identified through comparisons with-in cohorts at time of treatment, Table XIV columns four and eight, and with-in future classrooms models, Table XIV columns five and ten.

[22] Models account for all controls specified in equation (2).
[23] Estimates from the reading samples are qualitatively the same.

D.3 Do the Persistent Effects of NBPTS Teachers Differ by Grade Level?

Administrative education data has of recent become well detailed that even through high school grades you can capture with a high degree of precision all the teachers a student had. The increase in the quality of data has largely been due to the efforts of states to keep schools and teachers accountable. Due to the insurgence of better data for post primary school grades, recent studies (Goldhaber et al., 2015; Cavaluzzo, 2004), including Chapter II, have shown that NBPTS teachers impact students differentially by high school, middle school, and elementary school grades. The findings specifically show that the effect of NBPTS teachers on contemporaneous student achievement increases monotonically as we move up the grade distribution. The results suggest that the certification does a better job at identifying effective teaching in later grades.

However, effective teaching is also determined by a teacher's ability to improve longer-term outcomes. As such I test whether certified teachers continue to differentially impact student achievement one year out by middle and elementary school grades

D.3.1 Estimates Based on Classroom Variation

For models employing classroom variation I run, models similar to equation (1), however I interact all variables with an indicator for whether the student is in a middle school grade (6^{th} or 7^{th} grade). Results for math and reading can be found in Table XVII and Table XVIII respectively. The structure of these tables is similar to Table XIV with the exception that columns 1-5 provide estimates for elementary grades and columns 6-10 for middle school grades.

Across all specifications, I find evidence that certified teacher effects in elementary and middle school partially persist for mathematics. The estimates range between 0.007-0.016 of a standard deviation for elementary grades and between 0.011-0.029 of a standard deviation for middle school grades. Similar to the pattern found of contemporaneous effects by grade level, middle school NBPTS teachers' effects one year out are larger by 33%-50% relative to elementary grades. The smallest estimates for both middle school and elementary samples are from models accounting for the students' math teacher's productivity in *t+1*. Again, because treatment in t affects the quality of the student's next math teacher, the estimate from this model is subject to negative selection bias. Thus these estimates are lower bounds for the true persistent effects.

To understand the rate at which NBPTS teacher effects persistent by grade-level I run contemporaneous models on the analysis sample using the same specifications but on contemporaneous outcomes. These results are found in Panel B. of Tables XVII and XVIII. Comparing within specifications but across outcomes shows that the depreciation rate of the NBPTS effects for math in elementary grades is about 70%, which is similar to the depreciation rate for middle schools grades. For example, in column one of Table XVI the contemporaneous affect for math in elementary is 0.026 of a standard deviation. In Table XV, the persistence effect is 0.008 of a standard deviation, which is 70% smaller than the contemporaneous models.

The persistent effects for reading show a similar pattern. In fact across all but one specification in Panel A. of Table XVIII, the persistent effect of NBPTS teachers is 50% larger for middle school grades. The estimates range between 0.005-0.012 of a standard deviation for elementary grades and between 0.011-0.025 of a standard deviation. Comparing across contemporaneous models in Panel B. we see once again that reading effects persist close to fully one year out across both grade levels.

D.3.2 Estimates Based on Cohort Variation

Tables XIX and XX hold results from cohort level models for math and reading achievement outcomes respectively. Persistence estimates across the three models employed (school-by-grade fixed effects, school-by-year fixed effect, and both) show wildly differing estimates for both elementary and middle school samples. For math the estimates range from -0.026-0.008 of a standard deviation for elementary grades and -0.012-0.006 of standard deviation for middle school grades, and for reading the estimates range from 0 – 0.02 of a standard deviation for elementary grades and from -0.015-0.01 of standard deviation for middle school grade. All but one estimate is imprecisely estimated. Due to both the reduced variation and contamination bias prone to this estimation strategy, these estimates do not show a clear pattern. Contemporaneous models in Panel B. of both tables however do support that certified teachers are more effective.

E. Conclusion

Many studies show that NBPTS certified teachers improve student contemporaneous outcomes relative to non-certified teachers. These studies conclude that the NBPTS certification captures teacher quality, and can be used in policies used to identify/reward effective teaching. However, no study provides evidence that the certified teacher effects persist. This is an important question, because it has implications for how well the certification is a measure of teacher quality.

In this study I identify the effect of NBPTS certified teachers on student academic achievement in the following year. I exploit several sources of variation including across classroom variation within schools, within classroom variation, and across cohort variation. The analysis in this study reveals three important conclusions. First, this study demonstrates that NBPTS teacher effect do persist a year out. Preferred specifications show that NBPTS teachers continue to effect they're past students' academic achievement one year out by between 0.008-

0.015 of a standard deviation for math and between 0.007-0.012 of a standard deviation for reading. Second, the rate of persistence is higher for reading (90%) than math (35%). Lastly, the analysis reveals that special attention needs to be placed on empirical strategies that use cohort variation to study long-run effects when the probability of treatment varies overtime.

The first two results taken at face value suggest that contemporary NBPTS teacher effects are poor indicators for longer-term value added by these teachers particularly for math. Or alternatively, contemporary effects over-state the influence of NBPTS teachers on the permanent level of student knowledge. However persistence is not only driven by the effectiveness of teachers, but also by the extent the curriculum and subject matter tested on standardized tests is fluid over grades. In addition, it is also subject to the ability of students to retain information, and whether this knowledge influences true long-run outcomes such as wages. Interestingly, the rates of persistence between math and reading is quite stark even though many certified teachers who teach math also teach reading, particularly in elementary grades. Considering these differences it is likely that the curriculum is not necessarily as fluid for math relative to reading over grades. However it may also be the case that reading retention is much higher, because it is more likely to be reinforced outside of the classroom. While these factors likely explain part of the fade-out, it's unlikely they explain all of it.

Finally the varied econometric analyses employed in this study has led to an important methodological implication for longitudinal studies that use cohort variation to understand treatment impacts on long-term outcomes. This work needs to be cautious of treatment contamination of control cohorts particularly when treatment varies overtime and the cross-section variation is used. In this study, within schools certified teachers tend to be in few grades as there are few certified teachers to begin with. Cross-sectional variation across cohorts within these schools would drive comparisons between cohorts that are treated at different times, but with the outcome observed at the same time. This unfortunately produces a downward bias, even if treatment is randomly assigned. Importantly, estimation strategies that take into account the panel structure of the data may produce cleaner results than strategies that identify off aggregate cohort variation.

Table XII
Descriptive Statistics by Samples[a]

	Full Sample	Analysis Sample
Class Characteristics		
Class size	23.3	23.3
Teacher experience	11.72	11.73
Proportion with advanced degree	0.3	0.3
Proportion with NBPTS	0.12	0.12
Student Characteristics		
Math score	0.04	0.07
Reading score	0.03	0.06
Proportion economically disadvantaged	0.51	0.49
Proportion black	0.26	0.26
Proportion white	0.56	0.57
Proportion Hispanic	0.11	0.1
Proportion female	0.5	0.5
Proportion disabled	0.11	0.1
Proportion limited English proficient	0.05	0.05
Age	11.73	11.69
Observations	2266855	1688388

[a]Source Data: NCERDC

Table XIII
Demographic Adjusted Mean Differences of NBPTS Treated and Non-treated Students that Leave the Sample[a]

	Mean-Diff
Economically Disadvantage	-0.007**
	(0.002)
Black	0.000
	(0.002)
Hispanic	-0.002
	(0.001)
Other	0.003*
	(0.001)
Female	0.000
	(0.002)
Disabled	0.001
	(0.002)
Limited English proficiency	0.001
	(0.001)
Observations	2266855

[a]Source Data: NCERDC

Table XIV
Contemporaneous and Persistent Effects of NBPTS Teachers on Student Academic Achievement Using Classroom Variation[a,b]

	Math					Reading				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. ***Outcome in t+1***										
NBPTS Teacher in time t	0.009***	0.011***	0.013***	0.015***	0.008***	0.009***	0.011***	0.011***	0.012***	0.007***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Panel B. ***Outcome in t***										
NBPTS Teacher in time t	0.029***	0.034***	0.032***	0.032***	-	0.011***	0.012***	0.012***	0.013***	-
	(0.003)	(0.003)	(0.003)	(0.003)		(0.002)	(0.002)	(0.002)	(0.002)	
School-by-year fe	Yes	No	Yes	No	No	Yes	No	Yes	No	No
School-by-grade fe	No	Yes	Yes	No	No	No	Yes	Yes	No	No
School-by-grade-by-year fe	No	No	No	Yes	No	No	No	No	Yes	No
Teacher in t+1-by-year fe	No	No	No	No	Yes	No	No	No	No	Yes
Observations			1688388					1660326		

[a] All models include student controls consisting of age, indicators for race, gender, disability status, limited English proficiency indicator, economically disadvantaged and a cubic expansion in math and reading test scores taken in year prior to treatment; classroom characteristics consisting of mean subject-specific lagged test-score in t-1 and age, class-size, and proportion non-white, limited English proficient, disabled and economically disadvantaged; teacher characteristics consisting experience dummies and an indicator for advanced degree. All controls are measured at time of treatment except if otherwise specified. Models in columns 1-3, 5-8, and 10 additionally include grade-by-year dummies in year of treatment. Standard errors shown in parentheses are clustered at teacher-by-year level at time of treatment (* p<0.05 ** p<0.01 *** p<0.001).
[b]Source Data: NCERDC

Table XV
Contemporaneous and Persistent Effects of NBPTS Teachers on Student Achievement Using Cohort Variation[a,b]

	Math			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A.						
Test Outcome in t+1						
Proportion of students with NBPTS teachers in grade g	-0.007	0.000	0.002	-0.004	0.005	0.001
	(0.006)	(0.007)	(0.007)	(0.004)	(0.005)	(0.004)
Panel B.						
Test Outcome in t						
Proportion of students with NBPTS teachers in grade g	0.022***	0.048***	0.036***	0.004	0.011*	0.009*
	(0.006)	(0.007)	(0.007)	(0.004)	(0.005)	(0.004)
School-by-year fe	Yes	No	Yes	Yes	No	Yes
School-by-grade fe	No	Yes	Yes	No	Yes	Yes
Observations		1688388			1660326	

[a]All models include student controls consisting of age, indicators for race, gender, disability status, limited English proficiency indicator, economically disadvantaged, and a cubic expansion in math and reading test scores taken in year prior to treatment; grade characteristics consisting of mean subject-specific lagged test-score in t-1 and age, class-size, and proportion non-white, limited English proficient, disabled and economically disadvantaged; teacher characteristics aggregated to the school-by-grade-by-year level consisting of experience dummies and an indicator for advanced degree; grade-by-year dummies. All controls are measured at time of treatment except if otherwise specified. Standard errors shown in parentheses are clustered at school-by-grade-by-year level at time of treatment (* p<0.05 ** p<0.01 *** p<0.001).
[b]Source Data: NCERDC

Table XVI
Predicting the Probability of Having a NBPTS Teacher the Following Year with Whether the Student has One Currently[a,b,c]

	NBPTS in grade g+1 (1)	NBPTS in grade g+1 (2)	NBPTS in grade g+1 (3)	NBPTS in grade g+1 (4)
NBPTS teacher in grade g	-0.054***	0.008	-	-
	(0.004)	(0.009)		
Proportion of students with NBPTS teachers in grade g	-	-	-0.321***	0.009
			(0.011)	(0.009)
School-by-year fe	Yes	Yes	Yes	No
School-by-grade fe	No	No	No	Yes
Observations	1598992	214697[b]	1598992	1598992

Table XVII

Contemporaneous and Persistent Effects of NBPTS Teachers by Grade Level on Math Achievement Using Classroom Variation[a,b]

	Elementary School					Middle School				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A.										
Test Outcome in t+1										
NBPTS Teacher in time t	0.008*	0.014***	0.014***	0.016***	0.007**	0.018***	0.021***	0.011**	0.029***	0.017***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.004)	(0.004)	(0.005)	(0.003)	(0.003)
Panel B.										
Test Outcome in t										
NBPTS Teacher in time t	0.026***	0.032***	0.028***	0.029***	-	0.059***	0.069***	0.038***	0.065***	-
	(0.003)	(0.003)	(0.003)	(0.003)		(0.005)	(0.005)	(0.006)	(0.005)	
School-by-year fe	Yes	No	Yes	No	No	Yes	No	Yes	No	No
School-by-grade fe	No	Yes	Yes	No	No	No	Yes	Yes	No	No
School-by-grade-by-year fe	No	No	No	Yes	No	No	No	No	Yes	No
Teacher t+1-by-year fe	No	No	No	No	Yes	No	No	No	No	Yes
Observations			1688388					1660326		

[a]Each model is fully interacted with an indicator for whether the student is in middle school (grade 6 or 7). All models include student controls consisting of age, indicators for race, gender, disability status, limited English proficiency indicator, economically disadvantaged and a cubic expansion in math and reading test scores taken in year prior to treatment; classroom characteristics consisting of mean subject-specific lagged test-score in t-1 and age, class-size, and proportion non-white, limited English proficient, disabled and economically disadvantaged; teacher characteristics consisting experience dummies and an indicator for advanced degree. All controls are measured at time of treatment except if otherwise specified. Models in columns 1-3, 5-8, and 10 additionally include grade-by-year dummies in year of treatment. Standard errors shown in parentheses are clustered at teacher-by-year level at time of treatment (* p<0.05 ** p<0.01 *** p<0.001).

[b]Source Data: NCERDC

Table XVIII

Contemporaneous and Persistent Effects of NBPTS Teachers by Grade Level on Reading Achievement Using Classroom Variation[a]

	Elementary School					Middle School				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A.										
Test Outcome in t+1										
NBPTS Teacher in time t	0.009***	0.011***	0.012***	0.010***	0.005*	0.018***	0.022***	0.011***	0.025***	0.014***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Panel B.										
Test Outcome in t										
NBPTS Teacher in time t	0.008***	0.010***	0.012***	0.009***	-	0.023***	0.026***	0.014***	0.027***	-
	(0.002)	(0.003)	(0.002)	(0.002)		(0.003)	(0.003)	(0.004)	(0.003)	
School-by-year fe	Yes	No	Yes	No	No	Yes	No	Yes	No	No
School-by-grade fe	No	Yes	Yes	No	No	No	Yes	Yes	No	No
School-by-grade-by-year fe	No	No	No	Yes	No	No	No	No	Yes	No
Teacher t+1-by-year fe	No	No	No	No	Yes	No	No	No	No	Yes
Observations			1688388					1660326		

[a]Each model is fully interacted with an indicator if the student is in middle school (grade 6 or 7). All models include student controls consisting of age, indicators for race, gender, disability status, limited English proficiency indicator, economically disadvantaged and a cubic expansion in math and reading test scores taken in year prior to treatment; classroom characteristics consisting of mean subject-specific lagged test-score in t-1 and age, class-size, and proportion non-white, limited English proficient, disabled and economically disadvantaged; teacher characteristics consisting experience dummies and an indicator for advanced degree. All controls are measured at time of treatment except if otherwise specified. Models in columns 1-3, 5-8, and 10 additionally include grade-by-year dummies in year of treatment. Standard errors shown in parentheses are clustered at teacher-by-year level at time of treatment (* p<0.05 ** p<0.01 *** p<0.001).

[b]Source Data: NCERDC

Table XIX
Contemporaneous and Persistent Effects of NBPTS Teachers by Grade Level on Math Achievement Using Cohort Variation[a]

	Elementary School			Middle School		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A.						
Test Outcome in t+1						
Proportion of students with NBPTS teachers in grade g	0.008	-0.026**	-0.002	-0.003	-0.012	0.006
	(0.009)	(0.008)	(0.009)	(0.012)	(0.009)	(0.012)
Panel B.						
Test Outcome in t						
Proportion of students with NBPTS teachers in grade g	0.048***	0.013	0.020*	0.096***	0.049***	0.056*
	(0.009)	(0.008)	(0.009)	(0.012)	(0.009)	(0.015)
School-by-year fe	No	Yes	Yes	No	Yes	Yes
School-by-grade fe	Yes	No	Yes	Yes	No	Yes
Observations		1688388			1660326	

[a]Each model is fully interacted with an indicator for whether the student is in middle school (grade 6 or 7). All models include student controls consisting of age, indicators for race, gender, disability status, limited English proficiency indicator, economically disadvantaged, and a cubic expansion in math and reading test scores taken in year prior to treatment; grade characteristics consisting of mean subject-specific lagged test-score in t-1 and age, class-size, and proportion non-white, limited English proficient, disabled and economically disadvantaged; teacher characteristics aggregated to the school-by-grade-by-year level consisting of experience dummies and an indicator for advanced degree; grade-by-year dummies. All controls are measured at time of treatment except if otherwise specified. Standard errors shown in parentheses are clustered at school-by-grade-by-year level at time of treatment (* p<0.05 ** p<0.01 *** p<0.001).
[b]Source Data: NCERDC

Table XX

Contemporaneous and Persistent Effects of NBPTS Teachers by Grade Level on Reading Achievement Using Cohort Variation[a,b]

	Elementary School			Middle School		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A.						
Test Outcome in t+1						
Proportion of students with NBPTS teachers in grade g	0.012	0.000	0.020**	0.01	-0.005	-0.015
	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.009)
Panel B.						
Test Outcome in t						
Proportion of students with NBPTS teachers in grade g	0.013*	0.002	0.033***	0.021**	0.01	-0.01
	(0.006)	(0.006)	(0.006)	(0.017)	(0.006)	(0.009)
School-by-year fe	No	Yes	Yes	No	Yes	Yes
School-by-grade fe	Yes	No	Yes	Yes	No	Yes
Observations		1688388			1660326	

[a] Each model is fully interacted with an indicator for whether the student is in middle school (grade 6 or 7). All models include student controls consisting of age, indicators for race, gender, disability status, limited English proficiency indicator, economically disadvantaged, and a cubic expansion in math and reading test scores taken in year prior to treatment; grade characteristics consisting of mean subject-specific lagged test-score in t-1 and age, class-size, and proportion non-white, limited English proficient, disabled and economically disadvantaged; teacher characteristics aggregated to the school-by-grade-by-year level consisting of experience dummies and an indicator for advanced degree; grade-by-year dummies. All controls are measured at time of treatment except if otherwise specified. Standard errors shown in parentheses are clustered at school-by-grade-by-year level at time of treatment (* $p<0.05$ ** $p<0.01$ *** $p<0.001$).

b Source Data: NCERDC

IV. DIRUPTIVE PEERS AND THE ESTIMATION OF TEACER VALUE ADDED (Previously published as Horoi, Irina and Ben Ost. "Disruptive Peers and the Estimation of Teacher Value-Added." *Economics of Education Review*, 49, (2015):180-192.)

A. Introduction

Understanding classroom peer effects is important both for determining optimal student grouping patterns and for generally understanding the educational production function. While classroom peer effects have been studied extensively, most research has focused on how the existence or absence of peer effects influences whether students should be tracked or placed in heterogeneous classrooms. While these considerations are first order, the existence of peer effects also implies that the educational production functions typically estimated in the literature omit an important input. To the extent that these unmeasured peer inputs are correlated with other school and classroom inputs, estimates of non-peer inputs will be biased. This point is illustrated theoretically by Lazear (2001) in the context of estimating the returns to class size, but little research has examined how peer effects influence the estimated impact of non-peer inputs empirically.

In this study, we consider the extent to which peer effects bias the estimated impact of other inputs by showing how students who are likely to be disruptive influence the estimation of teacher value added. While teachers are just one input whose estimated impact could be biased by peer effects, the use of value-added estimates in high-stakes personnel decisions makes it particularly important to correctly estimate teachers' impacts.[24]

Many different forms of peer interactions have the potential to bias value-added estimation; we illustrate the issue in the context of disruptive students for several reasons. First, surveys of teachers and administrators frequently mention disruption as a major obstacle to

[24] As of 2013, 40 states require that a teacher's annual evaluation is based in part on her value added (Doherty and Jacobs 2013).

learning (Figlio, 2007). Research confirms that serious class disruption is a common occurrence, particularly in urban schools (Johnston, 2013; OECD, 2013). Second, while it is common for researchers to control for average peer demographic and peer academic performance when estimating teacher value added, it is rare to control for measures of disruption. Similarly, to the best of our knowledge, none of the value-added models currently in use to make high-stakes personnel decisions control for classroom disruption. Third, while there is a large literature on classroom peer effects, most of this research focuses on how peer academic performance impacts one's academic performance, and fewer studies explore how the non-cognitive attributes of one's peers impact one's academic performance.

While disruption is frequently reported as an issue by teachers and administrators, datasets typically do not include direct measures of disruption, so researchers necessarily use student characteristics that proxy for disruption (Carrell and Hoekstra, 2010; Fletcher, 2009[a], 2009[b]; Figlio, 2007; Friesen, Hickey, and Krauth, 2010). We follow this approach by using the diagnosis of an emotional disability to proxy for disruption. In the institutional context that we study, emotional disabilities are diagnosed primarily because students exhibit disruptive behaviors in school, and we show that emotional disability correlates strongly with disciplinary action such as suspension.[25] That said, without data on actual in-class behavior, it remains possible that the peer impacts that we document are due to some non-disruptive characteristic of emotionally disabled students.[26]

[25] Emotional disability is not a DSM medical diagnosis, but rather a designation used by schools to identify students in need of services due to their behavior. As such, if a student's disability does not manifest itself through school behavior, it will not be captured in our measure.

[26] While many students who are not ED transfer students may be disruptive, the extent of the disruption might differ between ED transfer students and other disruptive students. As such, we view our study as providing evidence that classroom disruption has the potential to meaningfully impact teacher value added, but we cannot provide empirical evidence as to the total impact of all forms of disruption on teacher evaluation.

This article expands the literature on classroom peer effects in several ways. First, we provide carefully identified evidence that peer non-cognitive attributes can influence academic achievement. Second, we use matched longitudinal data on students and teachers over a six-year period to show that the existence of these non-cognitive peer effects systematically influences the estimation of teacher value added. We show that for a variety of value-added models currently being used in policy, teaching emotionally disabled (ED) students reduces a teacher's estimated value added.

Identifying the impact of disruptive students on their peers is difficult because of the well-known issues of homophily, reflection and common shocks.[27] Our study addresses these concerns in several ways. First, we are able to address the possibility that students are non-randomly placed into classrooms by aggregating peer groups to the school-grade-year level and including a school-by-year fixed effect. Second, we focus on transfer students who were previously diagnosed as emotionally disabled to address concerns regarding reflection and common shocks (correlated effects). Finally, we test for non-random sorting into grades and find that the arrival of an emotionally disabled transfer student is uncorrelated with all observable predetermined characteristics, suggesting that homophily is unlikely to bias estimates of the peer effects we document.

Educational production functions invariably omit important inputs and we do not argue that this incompleteness necessarily leads to biased estimates of teacher quality. For example, parental and neighborhood inputs are rarely controlled for in value-added models, but since these

[27] "Homophily" refers to the idea that individuals may sort into groups based on their characteristics so that ego and peer outcomes will tend to be correlated in the absence of any peer effect. "Common shocks" refers to the idea that all of the individuals in a peer group may be exposed to the same inputs so that their outcomes will be correlated in the absence of any peer effect. "Reflection" refers to the difficulty between distinguishing the impact of peer characteristics on ego outcomes from the impact of ego characteristics on peer outcomes.

inputs are likely to be highly correlated over time, controlling for lagged test score or student fixed effects plausibly addresses many concerns regarding these omitted inputs.

Compared to omitting family or neighborhood characteristics, failing to control for peer effects presents a potentially more serious issue for value-added modeling for several reasons. First, since classmates change each year, peer effects will be time varying, and thus lagged test score will not control for current peer effects. Second, the majority of value-added models emphasize individual rather than peer controls, and these individual controls are unlikely to be good proxies for peer characteristics. While some researchers have controlled for average peer achievement and demographics when estimating teacher value added, few school districts collect or use data on peer quality in measuring teacher quality (Kane, 2014).

If disruptive students were randomly assigned to teachers, then the peer effects we document would make the estimation of yearly teacher value added more noisy, but these estimates would remain unbiased. Conversations with principals suggest, however, that the classroom placement of disruptive students is a non-random decision, and our data bear this out. We find that within a school-grade-year, emotionally disabled transfer students are nearly six percentage points more likely to be placed with male teachers. More broadly, we document non-random teacher assignments of many types of transfer students, providing clear evidence that the overall assignment of transfer students to teachers is not random. The systematic assignment of students to certain teachers may be optimal for student learning, but our study suggests that the practice imposes a cost on these teachers, particularly if value added is being used for high-stakes personnel decisions.

While our study is focused on teachers, the tension we highlight between worker evaluation and task assignment is applicable to a variety of occupations. For example, financial analysts are often times rewarded for accurate forecasts, but some analysts are assigned more difficult markets than others. Similarly, universities evaluate professors based on teaching evaluations, but the material in certain courses may be more easily accessible and appealing to students. Though pay-for-performance compensation schemes are theoretically effective at eliciting optimal effort, a critical difficulty in implementation is adjusting for task assignment difficulty. In contexts where identifying task difficulty is imperfect, randomly assigning tasks to workers ensures a more fair assessment of worker productivity, but may reduce total productivity by failing to capitalize on the comparative advantage of workers when assigning tasks. Pay-for-performance schemes that fail to adjust for task difficulty create perverse incentives in which workers with a comparative advantage in difficult tasks aim to hide this information from employers.

Relative to the evaluation systems in many other occupations, value-added models include substantial adjustment for task difficulty. Teachers who are assigned low-achieving students are not penalized simply because their students perform below average at the end of the year. That said, our study demonstrates that even value-added models are unable to fully adjust for task difficulty and as such, certain teachers are systematically misevaluated. In contexts where value-added models are used for high stakes teacher evaluation, there is an important balance to strike between fair assessment and optimal task assignment.

B. Related Literature

We focus on peer effects generated by emotionally disabled students, but classroom disruption is a broader phenomenon and a growing body of literature documents a relationship between a variety of measures of disruption and peer achievement. For example, Carrell and Hoekstra (2010) show that students from homes with domestic violence reduce the performance of their peers. Similarly, Aizer (2008) finds that unmedicated students with ADHD hurt the performance of their peers, but this effect dissipates once these students receive treatment. As in Aizer (2008), Figlio (2007) finds that proxies for disruption can be context dependent since he shows that boys with “girls’ names” hurt the performance of their peers in middle, but not elementary school.

There are several obstacles to studying the impact of disruption on peer outcomes. First, disruption, and its diagnosis, are likely endogenous to teacher and peer quality and thus it is necessary to use pre-determined or exogenous determinants of disruption to instrument or proxy for disruptive students. Second, as in all peer effects studies, it is necessary to address the possibility of common shocks (correlated effects) and homophily (sorting). These issues are further complicated by the fact that elementary students typically interact for many years in a row, so focusing on past performance or past behavior does not fully address issues of reflection or common shocks. For example, a chaotic environment might cause marginal students to misbehave and also reduce mean academic achievement. If the chaotic environment is caused by factors such as poor teacher quality, then this will lead to the appearance of peer effects, even if none exist.

The literature on non-cognitive peer effects uses a variety of empirical approaches, but the typical analysis uses within-school cohort variation in student characteristics to address issues of common shocks and homophily (Hoxby 2000; Carrell and Hoekstra 2010; Friesen, Hickey, and Krauth 2010; Lavy and Schlosser 2011; Neidell and Waldfogel 2010). Studies with access to only a single cohort of students instead rely on school and student fixed effects, and assume that there are no time-varying unobserved student characteristics that determine the placement of students into classrooms (Figlio 2007; Fletcher 2009[a], 2009[b]). Though findings regarding the impact of non-cognitive peer characteristics vary, studies that specifically consider disruption tend to find negative effects on peer performance.

Since Hoxby (2000) and Neidell and Waldfogel (2010) consider the impact of gender and race composition, reflection and common shocks are impossible in their context. However, studies that use potentially endogenous measures of peer behavior such as ED diagnosis or suspension rates require an instrument for disruption, and finding an appropriate instrument has proven difficult with standard data sources. Figlio (2007) and Carrell and Hoekstra (2010) are able to address the possibility that one's misbehavior may be caused by one's peers by collecting data to construct novel instruments. Figlio (2007) uses data on children's names to identify boys with traditionally girls' names as an instrument for disruptive behavior. He shows that these boys have a higher propensity to misbehave in middle school and their peers' test scores fall as a result. Carrell and Hoekstra (2010) use whether a student comes from a home with domestic violence as an instrument for disruptive behavior. They similarly find that disruptive students reduce the academic performance of their peers.

The study most similar to our own peer effects analysis is Kristoffersen et al. (2015), who consider the impact of three types of potentially disruptive students. Using Danish registry data, they identify students whose parents have criminal records, whose parents are divorced, and who have a psychiatric diagnosis. They then examine the impact of disruptive transfer students as measured by these three proxies for disruption. As in our own study, they define peer groups at the grade level to avoid issues of classroom sorting and they include school fixed effects to focus on across-cohort variation within a school. In their preferred specification, they find that transfer students with a preexisting psychiatric diagnosis reduce cohort level test scores by approximately 0.02 standard deviations for reading and 0.01 for math. Interestingly, although the population studied in their paper is quite different from the population we study, their estimates of the negative effects of disruptive students are quite similar to the estimates we present in Section VI.

We are aware of no study that specifically considers the impact of disruption on teacher value added, but several papers have explored the sensitivity of value-added estimates to controls for peer characteristics. Ballou et al. (2004) finds that value-added estimates are quite sensitive to the inclusion of peer characteristics, though they attribute this to a lack of variation in peer characteristics in their data, rather than an issue with the specification that omits peer characteristics. They note that their estimates of the impact of peer characteristics are implausibly large, and as such, they discount these specifications.[28]

In a working paper written in parallel with our own, Johnson, Lipscomb and Gill (2013) explores the importance of peer controls for value-added estimates and finds that value-added estimates are broadly robust to the inclusion of peer characteristics, though adding peer controls does result in a somewhat different ranking of teachers. Compared to our manuscript, their study is more focused on understanding how value-added estimates change when a large vector of peer characteristics are controlled for, as opposed to measuring the causal impact of any particular type of peer effect. As such, Johnson et al. (2013) take all peer characteristics as exogenous and do not attempt to address issues such as historical reflection, homophily, or common shocks.

Though not the focus of their paper, Aaronson, Barrow and Sander (2007) examine how including measures of peer absence rates and peer achievement influence teacher value-added estimates at the high school level. They find that controlling for neighborhood and student characteristics along with peer absence and lagged peer performance reduces the across-teacher standard deviation of the value-added estimates by approximately one third. Since their

[28] Ballou et al. (2004) notes that their estimates of the impact of peer poverty are likely biased, though they do not speculate as to why. One potential reason is that their model takes peer characteristics as exogenous and so the sorting of students to schools has the potential to bias estimates of the peer effect.

emphasis is on demonstrating that quality variation across teachers is not *entirely* explained by these factors, they do not specifically examine the extent to which peer controls impact teacher value added. Additionally, as they mention in the article, since teachers may influence the absence rate of their students, controlling for peer absence might be over-controlling.

C. Institutional Background

Serious emotional disturbance, or emotional disability, is one of the disabilities covered by the Individuals with Disabilities Education Act (IDEA), which governs how states provide interventions and services to disabled students. In North Carolina, students are identified as emotionally disabled by a team of evaluators including a regular teacher, a Special Education teacher, a licensed psychologist, a parent and a representative from the local education agency. Together they make the final decision on diagnosis that leads to an individualized education program. While students diagnosed with this disorder are a heterogeneous group, many of the behaviors typically used to diagnose an emotional disability can be directly linked to classroom disruption. For example, the screening and evaluation for emotional disability guidelines provided by the North Carolina Department of Public Instruction lists the following behaviors: "aggressive and authority challenging behaviors, overreaction to environmental stimuli, markedly diminished interest in activities, agitated, and physical manifestation of fear that have psychosomatic origin."[29]

Our data confirm the link between emotional disability and disruptive school behavior: students diagnosed with an emotional disability are 333 percent more likely to be suspended during sixth grade compared to other students. Relative to other proxies for disruption,

[29] This is not an exhaustive list of the behaviors that fit into one or more of the federally defined characteristics. For more examples and information see: http://ec.ncpublicschools.gov/instructional-resources/behavior-support/resources/screening-and-evaluation-for-serious-emotional-disability.

emotional disability is much more strongly related to school suspension. For example, males with female sounding names are 25 percent more likely to be suspended (Figlio, 2007) and exposure to domestic violence increases the number of disciplinary incidents by approximately 110 percent (Carrel and Hoekstra, 2010).

D. Data

This study uses restricted access student-teacher-matched data provided to us by the North Carolina Education Research Data Center (NCERDC). North Carolina's public school data contains rich information on students, classrooms, teachers, schools, and districts. It includes this information for all public school students in the state of North Carolina from 1995-2012. However, course membership information necessary for matching students accurately to classrooms is unavailable before 2006 and incomplete for that year. Therefore, we use data from 2007-2012 for the present analysis.

To create our estimation sample we start with student test score data from 2006-2012. We restrict it to fourth and fifth graders in the years 2007-2012, using 2006 to obtain baseline test scores. Additionally, we include only students who have taken mainstream standardized tests in math and reading and who have a baseline test score. Finally, to determine math and reading classrooms and their associated teachers, we use the 2007-2012 course membership data. These data allow us to match students to their official subject-specific classrooms and teachers. While our administrative data minimizes the extent of measurement error compared to a survey, there is likely some degree of measurement error in classroom assignments because students occasionally move between classrooms mid-year and some elementary schools use informal ability grouping that may not be reported in our data.

Table XXI shows the descriptive characteristics for the entire student sample, the sample of emotionally disabled (ED) students, and the ED transfer students that we use to identify disruption. ED transfer students have very different characteristics than the average student in our sample, and are somewhat lower performing than the average ED student. For example, compared to the average student, ED transfer students are thirty-four percentage points more likely to be male, twenty-eight percentage points more likely to be African American, and thirty percentage points more likely to come from an economically disadvantaged home. ED students also perform worse academically, scoring 0.98 standard deviations below average in math and 0.87 standard deviations below average in reading.[30] ED transfer students are more likely to be assigned black teachers and are somewhat more likely to be assigned teachers with an advanced degree.

While different from the average student, ED transfer students share more similarities with the average ED student, as can be seen by comparing columns three and four to five and six. For example, ED transfer students are only two percentage points more likely to be male, six percentage points more likely to be African American, perform only 0.1 standard deviations lower on their baseline math and reading tests, and tend to be placed in slightly smaller classrooms. That said, ED transfer students might be more disruptive than their observable characteristics would predict since they have been recently thrust into a new, unfamiliar environment and their school move likely occurred simultaneously with other shocks such as a residential move or parental divorce. The descriptive statistics in Table XXI suggest that ED transfer students are quite different from the average student and consequently might not be assigned to the same types of classrooms as the average student.

[30] Test scores have been normalized such that grade-by-year test score have mean zero with a standard deviation of one.

In order to directly examine the degree to which ED transfer students are non-randomly placed into classrooms, we examine the characteristics of classrooms with and without an ED transfer student. Comparing columns one and three in Table XXII shows that classrooms with an ED transfer student are considerably different than classrooms without an ED transfer student: about six percentage points more male, ten percentage points more African American, and eleven percentage point more economically disadvantaged.

While students in classrooms with an ED transfer student score 0.35 standard deviations lower on their math achievement test, it would be wrong to interpret this difference as the causal effect of the ED transfer student. This difference must be partly driven by student sorting since students who are in classrooms with an ED transfer student also scored 0.30 standard deviations worse on their math test in the previous year.

Simple empirical models using classroom variation will fail to control for the non-random selection of students in classrooms. Furthermore, even models that include school fixed effects may be biased since students can be sorted to classrooms within schools. For this reason we use grade variation in exposure to an ED transfer student combined with school-by-year fixed effects to identify estimates. In Section VI we test for and fail to find evidence that student sorting drives grade variation in ED transfers.

E. Empirical Strategy

E.1 Peer Effect on Student Achievement

Identifying and estimating the impact of high-needs children on their peers is complex due to issues of common shocks, reflection, and homophily (Manski, 1993; Moffitt, 2001). In our context, in order to identify the impact of peer effects, we must assume

1) Conditional on student and classroom controls, students who are exposed to an ED student are comparable to those who are not (no homophily).
2) Emotional disability diagnosis and peers' academic achievement are not simultaneously determined (no reflection).
3) There exist no unobservable inputs that determine both ED diagnosis and peer achievement (no common shocks).

Assumption (1) is unlikely to hold at the classroom level since Sections IV and Section VI show that classrooms with an ED transfer student have different observables both across and within schools. This evidence suggests that unless we have an instrument that creates random variation in classroom formation, classroom variation should not be used to identify the effect of exposure to an ED transfer student on student achievement. We address endogenous peer formation by defining peer groups at the grade level and including school-by-year fixed effects. Past studies have used this same aggregation strategy to address homophily (Friesen, Hickey, and Krauth, 2010; Carrel and Hoekstra, 2010; Hoxby, 2000). In Section VI we test directly for endogeneity of peers at the grade level, and find no evidence of it.

Assumption (2) requires that a student's emotional disability is not simultaneously determined with his peers' achievement. This assumption will be violated if peer achievement or behavior can cause a classroom environment in which marginal students are more likely to be diagnosed as emotionally disabled. To limit the possibility of reflection, we focus on transfer students who were diagnosed as emotionally disabled in their previous school, before being transferred. In our context, reflection is very improbable since new transfer students are unlikely to have ever had exposure to their current peers.[31]

[31] Student relocations have been previously used in the peer effect literature to resolve reflection (Imberman, Kugler, and Sacerdote, 2012).

Assumption (3) fails if common inputs cause both student performance and student behavior. For example, if an ineffective teacher contributes to ED diagnosis and reduced student performance, then it will appear that ED students reduce peer performance. For students who do not transfer schools, using a lagged measure of ED diagnosis only partly addresses this concern if achievement and diagnosis are persistent over time. The fact that we use past diagnosis of emotional disability of transfer students addresses this concern because transfer students do not share common inputs with their new peers until they arrive at their new school.

In light of the above issues, our preferred model is:

$$(1)\ Y_{igst} = \alpha GradeEDTS_{gst} + \omega Y_{igs(t-1)} + X_{it}\beta + G_{gst}\delta + \theta_{st} + \lambda_{gt} + \varepsilon_{igst}$$

where Y_{igst} is a subject-specific test score of individual *i* in grade *g* in school *s* at time *t*, $GradeEDTS_{gst}$ is an indicator set to 1 if the grade in school *s* at time *t* has an emotionally disabled transfer student, $Y_{igs(t-1)}$ is the lagged subject specific test score, X_{it} is a vector of student demographic information at time *t*, G_{gst} is a vector of grade-level peer characteristics of grade *g* in school *s* at time *t*, θ_{st} is a school-by-year fixed effect, λ_{gt} is a grade-by-year fixed effect, and ε_{igst} is an error term. Empirical specifications of education production functions with student lagged test scores on the right hand side are widely used because they are more flexible than gains models and can partly control for dynamic achievement-based sorting (Kane and Staiger 2008).

While equation (1) is our preferred specification, in the results section, we also show estimates that exclude the school fixed effects and the peer demographic and peer achievement controls to help describe the robustness of the result and establish the basic patterns in the data.

In all of our models we cluster the standard errors at the school-by-year-by-grade level, as this is the level of identifying variation.

E.2 Peer Effect on Teacher Value-Added

To evaluate how classroom disruption impacts estimated value added, we implement a two-step procedure. First, we estimate value added for every teacher in each year that they teach.[32] We allow value added to be time varying to mimic the value-added models typically used in policy. We then use these teacher-by-year value-added estimates as the dependent variable to assess whether teacher value added differs in years when a teacher teaches in a grade with an ED transfer student. We use grade, rather than classroom variation to address concerns that teachers are being sorted to particular classrooms within a grade (Rivkin et al., 2005).

Since different value-added models may yield different estimates, we consider three policy-relevant value-added models. The first model, which we refer to as the gains model, is very similar to the value-added models used in the Dallas DVASS model. In this model, we predict test score gains, adjusted for student-level covariates, to generate estimates of teacher-by-year value added. The second model is the value-added model currently in use by New York City and is also very similar to the model used by the Washington, DC public schools. This model controls for lagged test scores, student-level covariates and basic mean peer characteristics. The third model is based on one of the models used by the *Los Angeles Times* in their release of individual-level value-added estimates for teachers in the LAUSD. This model includes peer characteristics and student fixed effects in estimating teacher-by-year value added.

[32] Since many school districts only evaluate regular classroom teachers based on value added, we drop special education teachers from all analyses of teacher value added. Analyses that include special education teachers yield slightly larger point estimates and similar standard errors.

The three value-added models that we estimate are shown explicitly in equations (2a)-(2c)[33]:

$$(2a)\ \Delta Y_{ijcgst} = X_{it}\beta + \gamma_t + \mu_{jt} + \varepsilon_{icjgst}$$

$$(2b)\ Y_{ijcgst} = \omega Y_{ijcgs(t-1)} + X_{it}\beta + C_{jcst}\delta + \gamma_t + \theta_g + \mu_{jt} + \varepsilon_{icjgst}$$

$$(2c)\ \Delta Y_{ijcgst} = C_{jcst}\delta + \mu_{jt} + \pi_i + \varepsilon_{icjgst}$$

Y_{ijcgst} is subject-specific test score of individual *i* matched to subject-specific teacher *j* in classroom *c* in grade *g* in school *s* at time *t*, $Y_{ijcgs(t-1)}$ is lagged subject-specific test score of individual *i* matched to subject specific teacher *j* in classroom *c* in grade *g* in school *s*, X_{it} is a vector of student demographic information at time *t*, C_{jcst} is a vector of classroom-level characteristics of classroom *c* with teacher *j* in school *s* at time *t*, γ_t are year dummies, θ_g are grade dummies, μ_{jt} are teacher-by-year fixed effects, π_i is a student fixed effect and ε_{icjgst} is an error term. [34]

Using the teacher-by-year value-added estimates as the dependent variable, we examine the impact of exposure to an ED transfer student. Analogous to equation (1), we estimate (3):

[33] Student controls in (2a) and (2b) include dummies for male, race (African American, Hispanic, White), limited English proficient, economically disadvantaged, non-transfer emotionally disabled, emotionally disabled transfer, low-achieving transfer, other transfer, and other disability. Model (2b), in addition to the subject-specific lagged achievement, also includes other subject-lagged achievement. Classroom controls in models (2b) and (2c) include proportion male, African American, Hispanic, White, limited English proficient, economically disadvantaged, non-transfer emotionally disabled, and other disability, and class size and average lagged achievement in math and reading.

[34] Because classroom composition is perfectly collinear within a teacher-year, we estimate the classroom composition effect, using a three-step procedure as described by Isenberg and Walsh (2013) and used by NYC and DC. First we run specifications similar to (2b) and (2c), except that the teacher-by-year fixed effects are replaced with teacher-by-school fixed effects. This allows us to compare multiple classrooms for a teacher over time, which breaks the perfect collinearity between teacher and classroom composition. In step two, we use the estimated impact of classroom composition to calculate an adjusted subject-specific test score that nets out the contribution of the classroom characteristics. In the final step, we use the adjusted subject-specific scores in place of the actual test scores, and estimate (3), omitting classroom variables from the specification.

$$(3)\ \mu_{jst} = \alpha TeachGradeEDTS_{jst} + \theta_{js} + \varepsilon_{jst}$$

where μ_{jst} is the estimated subject-specific teacher effectiveness for teacher *j* in school *s* at time *t,* $\mathrm{TeachGradeEDTS_{jst}}$ is an indicator equal to 1 if the teacher *j* in school *s* at time *t* teaches in a grade with an ED transfer student. Depending on the specification, we also include a school fixed effect, a school-by-year fixed effect or a teacher-by-school fixed effect. All standard errors are clustered at the school-by-year-by-grade level.

F. Results

F.1 Peer Effect on Student Achievement

Table XXIII reports the effects of grade exposure to an ED transfer student on the math and reading achievement of other students. The first four columns show the effect for math achievement, and the last four columns report the estimates for reading achievement. Column one reports OLS estimates controlling for student-level demographics, lagged student achievement in math, and grade-by-year dummies.[35] In columns two and six, school fixed effects are added, in columns three and seven, school-by-year fixed effects are included in place of school fixed effects, and in columns four and right, grade-level peer characteristics are added.[36]

The results shown in Table XXIII provide evidence that emotionally disabled students impact the math scores of their peers. Estimates for math drop by 43 percent when school fixed effects are added (column two) but remain fairly stable (and statistically indistinguishable) when subsequently adding school-by-year and peer characteristics. The relative stability of the

[35] The student-level demographics include a male gender dummy, race dummies (African American, Hispanic, and White), a dummy for being economically disadvantaged, a dummy for limited English proficiency, a dummy for non-transfer emotionally disability, and a dummy for other disability.
[36] The grade-level peer controls include proportion male, proportion African American, proportion Hispanic, proportion White, proportion limited English proficiency, proportion economically disadvantaged, proportion non-transfer emotionally disabled, proportion other disability, and average subject-specific lagged achievement.

estimates in columns two through four is reassuring since it suggests that ED students do not systematically transfer to grades with different student characteristics. The fact that the estimated peer effect falls considerably when adding school fixed effects suggests that ED students tend to transfer to schools with lower-achieving students.

For reading, the estimates drop by more than half when school fixed effects are included (column 6), and stay statistically the same in magnitude, but become insignificant across more controlled specifications. In our preferred specification, shown in columns four and eight, a single emotionally disabled student causes the average performance of other students in the grade to be reduced by 0.017 standard deviations in math and 0.006 standard deviations in reading.[37]

To put our estimates in perspective, it is worth converting our estimates to approximate days of learning, as in Reardon (2011). Based on his estimates, average learning is approximately 0.3 standard deviations per year, and thus an effect of 0.017 corresponds to approximately a 6 percent reduction in school-year equivalents, or about two weeks less learning.

These findings are consistent with the literature on academic externalities associated with disruptive peers (Carrel and Hoekstra 2010; Figlio 2007; Friesen, Hickey, and Krauth 2010; Fletcher[a,b] 2009). That said, in comparing our estimates to some papers in the literature (e.g. Fletcher[b] 2009), it is important to keep in mind that we estimate our effects at the grade, rather than classroom level.[38] Our estimate is most directly comparable to Kristoffersen et al. (2015)

[37] Finding a smaller impact on reading scores is a consistent finding in the literature. The most likely explanation is that most math learning occurs at school, whereas reading is more likely to be learned both at school and at home.
[38] Estimating peer effects at the grade level is similar to using grade-level variation to instrument for having an ED transfer student in one's classroom. In fact, our preferred specification is simply the reduced form from that IV specification. To estimate the IV specification, one simply scales up our estimate of 0.017 based on the inverse probability that a student is in classroom with an ED transfer student, given that they are in a grade with an ED transfer student. We opt not to estimate the peer effects using the IV specification because this specification assumes

since they also aggregate peer groups to the grade level and examine the impact of transfer students with psychiatric diagnoses. Compared to Kristoffersen et al. (2015), our estimates are larger in math and smaller in reading, but the general range of estimates is similar.

To explore whether the impact of ED transfer students can be explained by their transfer status or by the fact that they are low-achieving, in Table XXIV we show how the estimated impact of ED transfer students changes when we control for the proportion of students transferring into a grade as well as whether there is a low-achieving transfer student entering the grade. For the reader's convenience, columns one and four of Table XXIV replicate the preferred estimates from columns four and eight of Table XXIII. Columns two and five show that controlling for the proportion of transfer students in the grade modestly reduces the ED transfer student coefficient. Columns three and six show that low-achieving transfer students do not appear to hurt their peer's performance as much as ED transfer students, suggesting that the ED transfer student effect is not simply due to the fact that ED transfer students tend to be low performing. In any case, for the analysis of teacher value added, it is unimportant whether classroom disruption or some other channel drives the peer effects, so long as a peer effect exists. Though not the focus of our analysis, it is interesting to note that consistent with the results in Hanushek, Kain and Rivkin (2004), Table XXIV shows that increasing the proportion of transfer students in general reduces student performance.

F.2 Specification Tests

To test for whether emotionally disabled transfer students endogenously enter particular grades in a school, we examine whether predetermined student characteristics predict whether or

---

that students have no interaction with students in their grade outside of their classroom, an assumption we find implausible. This is particularly implausible at the elementary level where students often mix across self-contained classrooms for special math or reading classes. In our data, we have no information regarding this sort of informal classroom assignment. That said, when the IV specification is used, our estimate is that ED transfer students reduce classroom peer performance by 0.101 standard deviations which is on par with the literature.

not one's school-grade-year has an ED transfer student. Specifically, we regress an indicator for whether or not the student is in a grade with an ED transfer student on student characteristics, school-by-year fixed effects and grade-by-year fixed effects. While our preferred estimates will only be biased if ED students sort to particular grades, we also examine whether ED students sort into particular classes within a school using the same approach. The idea behind these tests is to see whether ED transfer students enter grades or classes that were likely to perform poorly in any case.

These results are presented in Table XXV. Column 1 shows results for the classroom-level regression, and column two shows the results for the grade-level specification. We find that a number of predetermined student characteristics such as gender, Hispanic ethnicity, previous emotional disability, and pretest scores are predictive of the probability of being assigned a classroom with an ED transfer student within a school and year. For example, a student classified with an emotional disability in a previous year is 7.5 percentage points more likely to get assigned to a class with an ED transfer student. These results suggest that ED transfer students are systematically placed into certain classrooms and thus, across-class variation cannot be used to identify the impact of peer effects.

Column two of Table XXV shows that when we use grade variation within a school and year, we find that none of the student characteristics predict assignment to grades with an ED transfer student. This suggests that conditional on school-by-year fixed effects, ED transfer students are not systematically placed into grades with certain types of students.[39]

F.3 Teacher Value-Added

[39] Since not all students take both math and reading classes, there are slightly different samples for the math and the reading analyses. The specification tests shown in Table XXV use the sample of students in math classes but the results are qualitatively and quantitatively the same for the sample of students in reading classes. These results are available from the authors upon request.

Table XXVI presents the results from estimating equation (3): the effect of having an ED student in the grade on teacher value added. We measure ED exposure at the grade, rather than classroom level, since Table XXV shows that ED students are systematically sorted towards certain classes. In these specifications, we restrict the analysis to regular education teachers to ensure that special education teachers do not drive the results. This focus makes sense from a policy perspective since most special education teachers are not evaluated based on value added.[40]

Columns one through three show that regardless of the choice of value-added model, teachers are evaluated as less effective when they are in a grade with an ED student. Based on these results alone, however, it would be wrong to conclude that the ED students are causing the reduction in value added, because it is possible that ineffective teachers are more likely to work in schools that have more ED transfer students. To address this possibility, in columns four and five of Table XXVI, we use the same dependent variable as in column three, but we add school or school-by-year fixed effects to the model relating ED students to teacher value-added. These models use within-school variation in exposure to ED students to test whether teachers in grades with an ED student are evaluated as worse than other teachers in the same school.

Though student sorting to teachers within a grade cannot bias our estimates (since we aggregate to the school-grade-year level) it remains possible that lower-quality teachers are placed into grades that will have ED transfer students. We investigate this possibility by adding teacher-by-school fixed effects to the model that predicts teacher value added. Essentially, this specification uses across-time variation in exposure to ED transfer students to compare teachers to themselves. Column six of Table XXVI shows that teachers are evaluated as having lower value added (relative to themselves) in years that they are in a grade with an ED student. Past

[40] All estimates are nearly identical when we include special education teachers in the analysis.

research has found that there exists substantial within-teacher year-to-year variation in value-added estimates (McCaffrey et al., 2009). The result shown in column six suggests that a portion of this variation in teacher value added is attributable to peer effects.

In interpreting the magnitudes of the estimates shown in Table XXVI, it is important to keep in mind that the dependent variable is teacher value added—not student test scores. Our estimates imply that teacher value added for math is approximately 0.01 *student* standard deviations lower because of the ED student. Since the standard deviation of teacher value added is approximately one-fifth of the standard deviation of student test scores, our effect size corresponds to approximately 5 percent of a standard deviation decrease in teacher value added. The magnitude of our estimate suggests that few teachers will be grossly misevaluated as a result of the peer effects that we study, but the estimate is large enough to be of substantive significance. Also, we are identifying the impact of just one type of peer effect, so it remains possible that the overall importance of peers in the estimation of value added is substantial. Though a complete investigation of how all peer effects influence teacher value added is beyond the scope of this paper, Table XXVII shows that teachers also have lower estimated value added in years in which they are teaching a grade with a higher proportion of transfer students.

F.4 Systematic Placement of ED Students

If ED transfer students were randomly assigned to teachers, peer effects might cause biased assessments of teacher quality in particular years, but over the long run, no teacher would be systematically penalized. For a principal interested in maximizing the learning of her students, however, it makes little sense to randomly assign ED transfer students to teachers since certain teachers may be better equipped to handle these students. If some teachers are repeatedly assigned more difficult students and value-added estimates fail to account for disruption, certain teachers will have lower estimated value added, even when measured over many years. Importantly, the value-added measure for these teachers would accurately reflect the amount of learning that occurs in their classrooms, but it would not accurately reflect the teachers' skill or effort.

Past work has documented that students in general are systematically sorted to teachers (Clotfelter, Ladd and Vigdor, 2006; Rothstein, 2009; Kalogrides, Loeb and Béteille, 2013), but we are not aware of any work that studies the within-school sorting of transfer students.[41] We contribute to this literature on student sorting by investigating whether ED transfer students are plausibly randomly assigned to teachers within a school. In addition to specifically documenting the sorting patterns of ED transfer students, we investigate the sorting patterns of transfer students in general, as it is possible that transfer students are less likely to be sorted to particular teachers than non-transfer students since principals have more limited information about new transfer students when making classroom assignments.

To investigate whether certain types of transfer students are systematically assigned certain types of teachers, we use student characteristics to predict teacher characteristics, conditional on a school-by-year-by-grade fixed effect. If transfer students are randomly assigned to teachers within a school, we should find that none of the student characteristics predict any of the teacher characteristics. Specifically, we estimate a series of regressions where the dependent variable is a teacher characteristic and the independent variables are indicators for whether the student is a transfer student, an ED transfer student, a low-achieving transfer student, a black transfer student, a Hispanic transfer student, or a male transfer student. The coefficient on the transfer status indicator is thus an estimate of the sorting behavior of transfer students who are

[41] While our manuscript focuses on estimating the impact of ED transfer students, peer effects likely exist for all types of students, so the general sorting of students to teachers documented in past work suggests that peer effects may contribute to systematically biased estimates for certain teachers.

not ED, low achieving, black, Hispanic or male. We restrict the analysis to regular education teachers to ensure that results are not driven by special education teachers. Furthermore, for each regression, we restrict the analysis to school-year-grades that have variation across teachers for the particular characteristic. For example, when predicting whether students are assigned a male teacher, we restrict the sample to grades in which there is at least one male teacher and one female teacher, since in grades with no male teachers, the coefficient will be zero by construction. In interpreting the results of these regressions, we are focused on correlations rather than a causal interpretation since we are simply describing how a variety of factors correlate with student assignment.

Column one of Table XXVIII shows that emotionally disabled, low-achieving and male transfer students are all more likely to be assigned male teachers compared to non-transfer students. Column two shows that transfer students in general are assigned less experienced teachers and this effect is particularly large for low-achieving and minority students.[42] Columns three and four show evidence of racial and ethnic sorting such that black students are more likely to be assigned black teachers and Hispanic students are more likely to be assigned Hispanic teachers. Column five shows that low-achieving transfer students are slightly less likely to be assigned to teachers with a master's degree and male transfer students are slightly more likely to be assigned to teachers with a master's degree. Overall, Table XXVIII shows that the assignment of transfer students to teachers is clearly not random.

The non-random assignment of students to teachers documented in Table XXVIII suggests that the value-added estimates of some teachers may systematically be penalized, but a key question is whether this systematic sorting of students to teachers is efficient overall.

[42] Feng (2010) finds that more experienced teachers are less likely to be assigned special education students in general, but does not specifically consider transfer students. Feng's findings imply that early career teachers may be penalized if special education students impose negative peer effects on their classmates.

Systematic sorting has the potential to benefit students by more appropriately matching teacher skills to student needs. That said, systematic sorting could also be driven by advantaged students demanding the best overall teachers or the best teachers using political capital to obtain more desirable student assignments. If students are systematically sorted to teachers for reasons other than optimizing student achievement, then the practice may harm certain teachers with little benefit to student achievement. On the other hand, if students are systematically sorted to optimize student-teacher match quality, then even though systematic assignments may unfairly penalize certain teachers, the practice could improve overall efficiency.[43]

G. Conclusion

The landscape for high-stakes teacher evaluation policies has changed dramatically over the last five years. Since 2009, 25 states and the District of Columbia have adopted policies that require teacher evaluation to include objective measures of student achievement. More strikingly, the number of states that require student growth to be the major criterion in teacher evaluation increased by 500 percent, going from 4 to 20 states including D.C. in 2013 (Doherty and Jacobs, 2013). As evaluations of teachers continue to rely more heavily on teacher value-added estimates, it is important that policy makers are aware of the limitations and strengths of these estimates.

Given the difficulty of credibly identifying the impact of peer effects, we do not attempt to give a full characterization of how teacher value added is impacted by all types of peer effects. Instead, we show that a particular type of peer effect—namely the impact of emotionally

[43] A rigorous assessment of the overall benefits of systematic sorting is beyond the scope of this paper, since this analysis would need to incorporate two complex phenomena. First, it would need to incorporate a thorough understanding of which students *should* be assigned to which teachers. Second, it would need to incorporate a model of how bias in value-added estimates impacts teacher quality. This latter question is complex since biased value-added estimates can impact teacher quality by reducing the efficacy of incentive schemes based on value-added estimates and also by directly altering the composition of the teaching workforce in cases where value added is used for high-stakes decisions.

disabled students—moderately biases the evaluation of teacher value added. While we focus on providing empirical evidence for this peer effect, it is likely that other forms of peer effects also influence the estimation of teacher value added, such that the total bias caused by peer interactions could be quite large.

In a recent influential paper, Chetty, Friedman and Rockoff[a] (2014) provide evidence that value-added models yield approximately unbiased estimates of teacher quality and that these value-added estimates correlate with long-run student outcomes. Though these results have recently been challenged (Rothstein, 2014), a correlation between long-run outcomes and value added is consistent with our results for several reasons. First, our results imply that value-added estimates will be only modestly biased—well within the standard error of the Chetty et al.[a] (2014) estimates. Second, Chetty et al.[a] (2014) find relatively weak correlations across years within a teacher, allowing for the possibility that year-to-year variation is partly driven by factors such as changes in peer composition not captured by their controls. Finally, the correlation between long-run outcomes and teacher value added is completely consistent with the notion that students learn less when their peers are disruptive. The smaller learning gains made by these students could plausibly impact long-run outcomes, and as we show, teacher value added is reduced as well. In the presence of important peer effects, basic value-added models may still correctly identify student learning, but they do not necessarily identify teacher quality.

As school districts increasingly rely on value-added models for high-stakes personnel decisions, principals should be aware that these models do not fully adjust for classroom composition. Teachers that are consistently given difficult classrooms may be evaluated to be less effective than teachers given less difficult students, even if their true quality is equivalent. Male teachers may be at particular risk of being unfairly evaluated since we find that ED transfer students are disproportionately assigned to male teachers. Koedel (2016) suggests that value-added models can be modified to include a "proportionality" adjustment that ensures that poorer schools are not penalized for the students they teach. While theoretically a similar adjustment could be made at the teacher level to address unequal student assignment across teachers, these adjustments can only be implemented based on observable characteristics, and we suspect that most of the teacher characteristics principals use to determine the assignment of disruptive children are not measured in standard administrative data. That said, to the extent that disruptive students tend to be found at poorer schools, across-school proportionality adjustments can address differences in the consequences of peer effects across schools by ensuring that the distribution of teacher value added at poorer schools is similar to that at richer schools. Within schools, however, forcing that the teachers assigned disruptive students have similar value added to the teachers not assigned disruptive students is likely untenable because the underlying distribution of quality may differ substantially between these two groups, and much disruptive behavior is unobserved or endogenous to teacher practice.

As discussed in Lavy, Paserman, and Schlosser (2011), students may impact their peers through a variety of channels. While beyond the scope of the current paper, it would be interesting for future work to disentangle whether the peer effects we estimate are attributable to direct classroom disruption or to teachers altering their instruction (or time allocation) as a result of the ED student.

While our study demonstrates one limitation of value-added estimates, it is important to note that we provide little evidence on the question of whether school districts should use value added for high-stakes teacher evaluation. First, it is very possible that observation-based evaluations are also subject to bias from peer effects. Though observers aim to evaluate teacher quality, observer perception of quality may be influenced by classroom composition (Whitehurst, Chingos and Lindquist 2014). Second, the magnitude of the bias we document is sufficiently modest so that the cost of unfairly evaluating some teachers may be outweighed by other benefits of value-added evaluation. Finally, regardless of any limitations in the estimation of teacher value added, policies that evaluate teachers based on value added may induce effort that improves student achievement (Dee and Wyckoff, 2013).

Table XXI
Descriptive Statistics[a]

	All Students		*Emotionally Disabled Students*		*Emotionally Disabled Transfer Students*	
	mean	sd	mean	sd	mean	sd
Male	0.50	0.50	0.82	0.39	0.84	0.37
African American	0.26	0.44	0.48	0.50	0.54	0.5
Hispanic	0.12	0.32	0.02	0.15	0.03	0.16
White	0.54	0.50	0.43	0.49	0.35	0.48
Limited English proficiency	0.07	0.25	0.01	0.11	0.02	0.13
Economically disadvantaged	0.51	0.50	0.76	0.43	0.81	0.39
Suspended in 6th grade	0.12	0.32	0.52	0.5	0.54	0.5
Reading score	0.00	1.00	-0.76	1.02	-0.87	0.97
Math score	0.00	1.00	-0.83	0.99	-0.98	0.93
Reading pretest	0.04	0.97	-0.70	1.00	-0.8	0.96
Math pretest	0.04	0.97	-0.75	0.95	-0.85	0.89
Male teacher	0.08	0.28	0.12	0.32	0.12	0.33
African American teacher	0.2	0.4	0.31	0.46	0.37	0.48
Hispanic teacher	0.01	0.08	0.01	0.07	0.004	0.06
White teacher	0.78	0.42	0.68	0.47	0.62	0.49
Teacher experience	11.34	9.16	11.34	9.00	10.86	8.83
Teacher with advanced degree	0.3	0.46	0.33	0.47	0.35	0.48
Observations	1311480		3902		1128	

[a]Source Data: NCERDC

Table XXII
Descriptive Statistics of Classrooms With and Without an Emotionally Disabled Transfer Student[a]

	Classroom without EDTS		*Classroom with EDTS*	
	mean	sd	mean	sd
Proportion male	0.51	0.14	0.57	0.18
Proportion African American	0.28	0.27	0.40	0.31
Proportion Hispanic	0.12	0.15	0.11	0.13
Proportion White	0.53	0.31	0.42	0.32
Proportion limited English proficiency	0.07	0.12	0.07	0.10
Proportion economically disadvantaged	0.54	0.28	0.66	0.27
Teacher experience	11.59	9.20	10.92	8.69
Class size	22.65	11.48	20.39	12.77
Math pretest score	-0.02	0.61	-0.32	0.63
Reading pretest score	-0.02	0.59	-0.29	0.62
Math score	-0.07	0.64	-0.42	0.68
Reading score	-0.07	0.61	-0.39	0.66
Observations	59359		677	

[a]Source Data: NCERDC

Table XXIII
Estimates of Exposure to an Emotionally Disabled Transfer Student on Math and Reading Test Achievement[a,b]

	Math				Reading			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Grade has an EDTS	-0.0339***	-0.0192***	-0.0157**	-0.0168***	-0.0237***	-0.0098**	-0.0062	-0.0064
	(0.0064)	(0.0056)	(0.0062)	(0.0061)	(0.0044)	(0.0039)	(0.0039)	(0.0039)
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade controls	No	No	No	Yes	No	No	No	Yes
School FE	No	Yes	No	No	No	Yes	No	No
School-by-year FE	No	No	Yes	Yes	No	No	Yes	Yes
Observations	1164306	1164306	1164306	1164306	1156700	1156700	1156700	1156700

Notes: [a]All models include student controls (male, African American, Hispanic, White, limited English proficiency, economic disadvantage, non-transfer emotionally disabled, other disability, and the subject specific pretest score), and grade-by-year dummies. Grade controls refer to proportion male, African American, Hispanic, White, limited English proficiency, economically disadvantaged, non-transfer emotionally disabled, other disability, and average subject specific lagged achievement. Standard errors clustered at the school-by-year-by-grade level are shown in parentheses (* p<0.10 ** p<0.05 *** p<0.01).
[b]Source Data: NCERDC

Table XXIV

Estimates of Exposure to an Emotionally Disabled Transfer Student and Other Transfer Students on Math and Reading Achievement[a,b]

	Math (1)	Math (2)	Math (3)	Reading (4)	Reading (5)	Reading (6)
Grade has an EDTS	-0.0168***	-0.0143**	-0.0144**	-0.0064	-0.0047	-0.0047
	(0.0061)	(0.0060)	(0.0060)	(0.0039)	(0.0039)	(0.0039)
Proportion other transfer students		-0.0434***	-0.0422***		-0.0386***	-0.0382***
		(0.0133)	(0.0133)		(0.0125)	(0.0125)
Grade has low achieving transfer student			-0.0059			-0.0018
			(0.0039)			(0.0024)
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
Grade controls	Yes	Yes	Yes	Yes	Yes	Yes
School-by-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1164306	1164306	1164306	1156700	1156700	1156700

Notes: [a]All models include student controls (male, African American, Hispanic, White, limited English proficiency, economic disadvantage, non-transfer emotionally disabled, other disability, and the subject specific pretest score), and grade-by-year dummies. Grade controls refer to proportion male, African American, Hispanic, White, limited English proficiency, economically disadvantaged, non-transfer emotionally disabled, other disability, and average subject specific lagged achievement. Standard errors clustered at the school-by-year-by-grade level are shown in parentheses (* p<0.10 ** p<0.05 *** p<0.01).

[b]Source Data: NCERDC

Table XXV
Predicting the Probability of Exposure to an Emotionally Disabled Transfer Student[a,b]

	Estimated at Class (1)	*Estimated at Grade* (2)
Male	0.0006***	-0.0001
	(0.0001)	(0.0002)
African American	-0.0002	-0.0006
	(0.0004)	(0.0005)
Hispanic	-0.0010**	0.0004
	(0.0005)	(0.0006)
White	-0.0002	-0.0003
	(0.0004)	(0.0004)
Previous emotional disability	0.0753***	-0.0035
	(0.0073)	(0.0030)
Math pretest	-0.0009***	0.0000
	(0.0002)	(0.0003)
Reading pretest	-0.0004**	-0.0001
	(0.0002)	(0.0002)
Grade has an emotionally disabled transfer student	0.1677***	-
	(0.0044)	
School-by-year FE	Yes	Yes
Observations	1154589	1154589

Notes: [a]All models include grade-by-year dummies. The samples refer to students in math classes. Standard errors clustered at the school-by-year-by-grade level are shown in parentheses (* p<0.10 ** p<0.05 *** p<0.01).
[b]Source Data: NCERDC

Table XXVI
The Relation Between Estimates of Teacher Quality and Whether or Not the Teacher Teaches in a Grade with an Emotionally Disabled Transfer Student[a,b]

Model used to calculate value added:	Gains	Student fixed-effect	Lagged achievement			
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Math Teachers						
Teaches math in grade with an emotionally disabled transfer student	-0.0229***	-0.0296**	-0.0257***	-0.0137**	-0.0134*	-0.0092**
	(0.0071)	(0.0141)	(0.0068)	(0.0061)	(0.0069)	(0.0046)
School FE	No	No	No	Yes	No	No
School-by-year FE	No	No	No	No	Yes	No
Teacher-by-school FE	No	No	No	No	No	Yes
Observations	45904	45400	45398	45398	45398	45398
Panel B. Reading Teachers						
Teaches reading in grade with an emotionally disabled transfer student	-0.0138***	-0.0319**	-0.0205***	-0.0096**	-0.0050	-0.0075*
	(0.0049)	(0.0127)	(0.0047)	(0.0043)	(0.0046)	(0.0040)
School FE	No	No	No	Yes	No	No
School-by-year FE	No	No	No	No	Yes	No
Teacher-by-school FE	No	No	No	No	No	Yes
Observations	49862	49297	49291	49291	49291	49291

Notes: [a]Panel A. shows results from regressions of math teacher value-added on a dummy for whether a teacher teaches math in a grade with an emotionally disabled student. Panel B. estimates the same specifications but for reading teachers and their value-added. In column 1, the dependent variable is value-added calculated using the gains model. In column 2, the dependent variable is value-added calculated using the student fixed effects model and in columns 3-6, the dependent variable is value-added calculated using the lagged achievement model (see text for details on the calculation of value added for each of these models). Standard errors clustered at the school-by-year-by-grade level are shown in parentheses (* $p<0.10$ ** $p<0.05$ *** $p<0.01$).
[b]Source Data: NCERDC

Table XXVII
The Relation Between Estimates of Teacher Quality and Whether or Not Teacher Teaches in a Grade with an Emotionally Disabled Transfer Student Conditional on Peer Characteristics of Transfer Students[a,b]

	(1)	(2)	(3)
Panel A. Math Teachers			
Teaches math in grade with an emotionally disabled transfer student	-0.0092**	-0.0075	-0.0077*
	(0.0046)	(0.0046)	(0.0046)
Proportion transfer in grade math teacher teaches in		-0.0395***	-0.0370***
		(0.0096)	(0.0097)
Teaches math in grade with low achieving transfer student			-0.0041
			(0.0028)
Teacher-by-school FE	Yes	Yes	Yes
Observations	45398	45398	45398
Panel B. Reading Teachers			
Teaches reading in grade with an emotionally disabled transfer student	-0.0075*	-0.0058	-0.0059
	(0.0040)	(0.0040)	(0.0040)
Proportion transfer in grade reading teacher teaches in		-0.0433***	-0.0416***
		(0.0076)	(0.0077)
Teaches reading in grade with low achieving transfer student			-0.0026
			(0.0022)
Teacher-by-school FE	Yes	Yes	Yes
Observations	49291	49291	49291

Notes: [a]Panel A. shows results from regressions of math teacher value-added on a dummy for whether a teacher teaches math in a grade with an emotionally disabled student. Panel B. estimates the same specifications but for reading teachers and their value-added. All teacher value-added measures were estimated with a lagged achievement specification which includes student controls (lagged achievement in reading and math and dummy indicators for male, African American, Hispanic, White, limited English proficient, economically disadvantaged, emotionally disabled transfer, emotionally disabled non-transfer, low-achieving transfer, other transfer, other disability), classroom controls (class size, classroom averages in lagged math achievement and lagged reading achievement, and proportion African American, Hispanic, White, limited English proficient, economically disadvantaged, non-transfer emotionally disabled, other disability) and teacher-by-year fixed effects. The model is the same as the one used in columns 3-6 of Table XXVI (see text for additional details on the calculation of value added). Standard errors clustered at the school-by-year-by-grade level are shown in parentheses. (* p<0.10 ** p<0.05 *** p<0.01).
[b]Source Data: NCERDC

Table XXVIII
Evidence of Sorting of Transfer Students to Teachers[a,b]

	Male Teacher	Experience	Black Teacher	Hispanic Teacher	Teacher with Advanced Degree
	(1)	(2)	(3)	(4)	(5)
Transfer student	-0.0058	-0.0829*	-0.0084*	-0.0564***	-0.0044
	(0.0042)	(0.0488)	(0.0048)	(0.0142)	(0.0036)
Emotionally disabled transfer student	0.0555*	-0.1404	-0.0104	-0.0865	0.0186
	(0.0334)	(0.3606)	(0.0323)	(0.1118)	(0.0297)
Low achieving math transfer student	0.0168***	-0.1518**	-0.0067	0.0369**	-0.0086*
	(0.0054)	(0.0607)	(0.0051)	(0.0172)	(0.0045)
Hispanic transfer student	0.0056	-0.2174***	0.0042	0.0823***	-0.0082
	(0.0068)	(0.0740)	(0.0066)	(0.0225)	(0.0055)
African American transfer student	-0.0055	-0.0995*	0.0298***	-0.0197	-0.0041
	(0.0046)	(0.0544)	(0.0048)	(0.0190)	(0.0040)
Male transfer student	0.0183***	-0.0003	0.0062*	0.0260**	0.0054*
	(0.0036)	(0.0405)	(0.0035)	(0.0132)	(0.0031)
School-by-grade-by-year FE	Yes	Yes	Yes	Yes	Yes
Observations	353881	820967	317996	25717	571108

Notes: [a]Columns 1 and 3-5 hold results from linear probability regression models, and columns 2 holds results from a linear regression. The models in each column are run on the sample of students in grades with variation in the dependent variable. Standard errors clustered at the school-by-year-by-grade level are shown in parentheses. (* p<0.10, ** p<0.05, *** p<0.01)
[b]Source Data: NCERDC

V. CONCLUSION

A. Synthesis

In this thesis I provide evidence that both teacher and peer quality are important inputs for the production of student learning. In addition, my studies highlight the importance of accounting simultaneously for these inputs when estimating the education production function to be able to parse out their individual impacts.

Chapter II and III show that students exposed to NBPTS teachers outperform their peers both contemporaneous and in future periods all else held fixed. These studies suggest that NBPTS certification is a measure of teacher quality that picks up fixed differences in quality rather than inducing changes in teachers' quality. This is particularly true for middle school teachers, where advanced subject expertise plays a stronger role in determining the quality of instruction.

Chapter IV on the other hand highlights a negative classroom externality, disruptive students and their impact on their peers' academic achievement. As expected we find that disruptive students negatively impact their peers' academic achievement in a non-trivial way. Because peer effects are rarely accounted for in the estimation of teacher value added, we show that teacher value added estimates are negatively impacted by disruptive peers. Importantly, we show that the assignment of disruptive students is non-random, so these peer effects lead to biases that do not impact the evaluation of every teacher equally. While we focus on providing empirical evidence for this peer effect, it is likely that other forms of peer effects also influence the estimation of teacher value added, such that the total bias caused by peer interactions could be quite large.

B. Implications for Policy and Future Research

As other studies have highlighted peer and teacher quality are two important inputs in consideration for policies that aim at improving student outcomes. Recent literature, including Chetty et al.[b] (2014) emphasize that identifying and retaining good teachers is an economical way of raising the quality of instruction. Peer effects on the other hand effect the efficacy of policies such as school choice, academic tracking, and optimal student sorting. Further, they have implication for the estimation of other school inputs. This is particularly true for teacher value added, which is used in high stakes personnel decisions in over 40 states.

Chapter II and III speaks directly to that of identifying and retaining good teachers. Taken together they suggest that the National Boards could be used for identifying good teachers when hiring, particularly when there are little to no meaningful objective measures of their quality. To date we know that traditional measures used for screening and salary schedules such as Advanced Degree, traditional teaching certificates, and teacher experience have little to no impact on student learning. Further, subjective evaluations such as those conducted by principals have been known to be compressed and thus provides little information. Lastly, teacher value added measures while informative, could be noisy if only one year of data is used on the teacher. Moreover, many subject specific teachers like History or Science teachers will not have a value added score because States and Districts do not test on these subjects. Considering these contexts, the National Board certification has an important role to play in improving the quality of instruction within schools as it provides an objective evaluation that is related to improved student outcomes.

Future work on the National Boards should focus on longer term outcomes such as wages. This will not only test the efficacy of national board teachers on long-run student outcomes, but it will also enable researchers to elaborate on the cost effectiveness of using NBPTS for hiring and retention policies. While Chapter II attempts to identify the present value lifetime earnings gains to a typical class from having a NBPTS teacher and relate it to the full costs of the certification, it's limitations are driven by extrapolating estimates of earning gains due to movements up the teacher value added distribution in New York City (Chetty et al.[b], 2014), and assuming that the NBPTS effect persists fully in future period. Both of these assumptions likely overstate the value of an NBPTS teacher.[44]

On the contrary Chapter IV relates to policies surrounding whether students should be tracked or placed in heterogeneous classrooms, and how much emphasis should be placed on teacher value added measures when evaluating teachers considering the biases we characterize. While we do find that disruptive students negatively impact their peers' academic achievement, our analysis is limited in understanding the full extent of peer interactions. If peer effects are on average positive such that bad peers gain more from being with good peers than good peers loose from being with bad peers, than optimal sorting would suggest a heterogeneous mixture. Unfortunately, our current analysis is unable to speak to this.

What we do know is that the disruptive students we identify our impact from are non-randomly sorted to teachers. This not only documents that administrators feel that some teachers may be better prepared to handle these students, but also that these peer effects if unaccounted for bias the estimation of teacher value added. Considering the large emphasis now placed on teacher value added in high stakes personnel decisions at the State and District level, principals should be aware that these models do not fully adjust for classroom composition. Teachers that are consistently given difficult classrooms may be evaluated to be less effective than teachers given less difficult students, even if their true quality is equivalent. However, it is also important

[44] Chapter III provides preliminary evidence that NBPTS teachers effects do not persist fully in future periods.

to highlight that observation-based evaluations are also subject to bias from peer effects. Though observers aim to evaluate teacher quality, observer perception of quality may be influenced by classroom composition (Whitehurst, Chingos and Lindquist 2014). Further, regardless of any limitations in the estimation of teacher value added, policies that evaluate teachers based on value added may induce effort that improves student achievement (Dee and Wyckoff 2013).

APPENDICIES

APPENDIX A. EVIDENCE OF IRB APPROVAL

Chapter II & III:

UNIVERSITY OF ILLINOIS
AT CHICAGO

Office for the Protection of Research Subjects (OPRS)
Office of the Vice Chancellor for Research (MC 672)
203 Administrative Office Building
1737 West Polk Street
Chicago, Illinois 60612-7227

Approval Notice

Initial Review (Response To Modifications)

August 22, 2014

Irina Horoi

Economics

601 S Morgan, 750 University Hall

M/C 144

Chicago, IL 60607

Phone: (517) 505-4579

RE: Protocol # 2014-0471

"Teachers, Peers, and Student Achievement"

Dear Ms. Horoi:

APPENDIX A. EVIDENCE OF IRB APPROVAL (continued)

Please note that UIC IRB approval is predicated upon matching of the terms of the data transfer agreement with the research terms and procedures contained in the investigator's application. Kindly submit a finalized (signed) copy of the data transfer agreement from the North Carolina Education Research Data Center, accompanied by an Amendment form, to obtain finalized approval from the UIC IRB.

Your Initial Review (Response To Modifications) was reviewed and approved by the Expedited review process on August 14, 2014. You may now begin your research

Please note the following information about your approved research protocol:

Protocol Approval Period: August 14, 2014 - August 14, 2015

Approved Subject Enrollment #: 20000000

Additional Determinations for Research Involving Minors: The Board determined that this research satisfies 45CFR46.404 ", research not involving greater than minimal risk. Therefore, in accordance with 45CFR46.408 ", the IRB determined that only one parent's/legal guardian's permission/signature is needed. Wards of the State may not be enrolled unless the IRB grants specific approval and assures inclusion of additional protections in the research required under 45CFR46.409 '. If you wish to enroll Wards of the State contact OPRS and refer to the tip sheet.

Performance Sites: UIC

Sponsor: None

Research Protocol(s):

a) Teachers, Peers, and Achievement; Version 1; 05/05/2014

Recruitment Material(s):

a) No recruitment materials will be used- analysis of secondary data obtained through an Agreement with the North Carolina Education Research Data Center.

Informed Consent(s):

a) Waiver of informed consent granted [45 CFR 46.116(d)] for the analysis of secondary data obtained through an Agreement with the North Carolina Education Research Data Center; minimal risk.

Your research meets the criteria for expedited review as defined in 45 CFR 46.110(b)(1) under the following specific category(ies):

(5) Research involving materials (data, documents, records, or specimens) that have been collected, or will be collected solely for nonresearch purposes (such as medical treatment or diagnosis)., (7) Research

on individual or group characteristics or behavior (including but not limited to research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

Please note the Review History of this submission:

Receipt Date	Submission Type	Review Process	Review Date	Review Action
05/14/2014	Initial Review	Expedited	05/19/2014	Modifications Required
06/13/2014	Response To Modifications	Expedited	06/19/2014	Modifications Required
08/12/2014	Response To Modifications	Expedited	08/14/2014	Approved

Please remember to:

→ Use your **research protocol number** (2014-0471) on any documents or correspondence with the IRB concerning your research protocol.

→ Review and comply with all requirements on the enclosure,

"UIC Investigator Responsibilities, Protection of Human Research Subjects"
(http://tigger.uic.edu/depts/ovcr/research/protocolreview/irb/policies/0924.pdf)

Please note that the UIC IRB has the prerogative and authority to ask further questions, seek additional information, require further modifications, or monitor the conduct of your research and the consent process.

Please be aware that if the scope of work in the grant/project changes, the protocol must be amended and approved by the UIC IRB before the initiation of the change.

We wish you the best as you conduct your research. If you have any questions or need further help, please contact OPRS at (312) 996-1711 or me at (312) 355-0816. Please send any correspondence about this protocol to OPRS at 203 AOB, M/C 672.

APPENDIX A. EVIDENCE OF IRB APPROVAL (continued)

Sincerely,

Alison Santiago, MSW, MJ

IRB Coordinator, IRB # 2

Office for the Protection of Research Subjects

cc: Steven G. Rivkin, Economics, M/C 144
Darren Lubotsky (Faculty Sponsor), Economics, M/C 144

Chapter IV:

UNIVERSITY OF ILLINOIS
AT CHICAGO

Office for the Protection of Research Subjects (OPRS)
Office of the Vice Chancellor for Research (MC 672)
203 Administrative Office Building
1737 West Polk Street
Chicago, Illinois 60612-7227

Approval Notice

Initial Review (Response To Modifications)

December 3, 2012

Ben Ost

Economics

601 S Morgan, 718 University Hall

M/C 144

Phone: (617) 233-3304

RE: **Protocol # 2012-1007**

"The Impact of Special-Education Schools on High-Need Students and their Peers"

Dear Dr. Ost:

Your Initial Review application (Response To Modifications) was reviewed and approved by the Expedited review process on November 30, 2012. You may now begin your research.

Please note the following information about your approved research protocol:

APPENDIX A. EVIDENCE OF IRB APPROVAL (continued)

Please remember to submit a copy of the completed data transfer agreement with the North Carolina Education Research Data Center. Data transfer agreements usually originate with the organization allowing access to the data and are counter-signed/completed by the UIC ORS (Office of Research Services, grants and contracts). **A copy of the completed agreement must be accompanied by an Amendment form when submitted to the UIC IRB.**

Please note that Jin Man Lee cannot be added as key research personnel at this time as he has no investigator training on file at UIC and is not currently eligible to engage in research protocols or to have access to identifiable data at UIC. All investigators and key research personnel involved in human subjects research must complete a minimum of two hours of investigator training in human subjects protection every two years.

Protocol Approval Period: November 30, 2012 - November 30, 2013

Approved Subject Enrollment #: 20,000,000 cases

Additional Determinations for Research Involving Minors: The Board determined that this research satisfies 45CFR46.404, research not involving greater than minimal risk.

Performance Site: UIC

Sponsor: None

Research Protocol:

b) The Impact of Special Education Schools on High-Need Students and Their Peers; Version 1; 12/14/2012

Recruitment Material:

b) Analysis of secondary data that will be obtained under a data use agreement - no recruitment materials used

Informed Consent:

b) Waivers of assent and permission have been granted under 45 CFR 46.116(d) for data that will be obtained under a data use agreement (minimal risk)

c) Your research meets the criteria for expedited review as defined in 45 CFR 46.110(b)(1) under the following specific category:

APPENDIX A. EVIDENCE OF IRB APPROVAL (continued)

(7) Research on individual or group characteristics or behavior (including but not limited to research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

Please note the Review History of this submission:

Receipt Date	Submission Type	Review Process	Review Date	Review Action
11/15/2012	Initial Review	Expedited	11/19/2012	Modifications Required
11/29/2012	Response To Modifications	Expedited	11/30/2012	Approved

Please remember to:

→ Use your **research protocol number** (2012-1007) on any documents or correspondence with the IRB concerning your research protocol.

→ Review and comply with all requirements on the enclosure,
"**UIC Investigator Responsibilities, Protection of Human Research Subjects**"

Please note that the UIC IRB has the prerogative and authority to ask further questions, seek additional information, require further modifications, or monitor the conduct of your research and the consent process.

Please be aware that if the scope of work in the grant/project changes, the protocol must be amended and approved by the UIC IRB before the initiation of the change.

We wish you the best as you conduct your research. If you have any questions or need further help, please contact OPRS at (312) 996-1711 or me at (312) 996-2014. Please send any correspondence about this protocol to OPRS at 203 AOB, M/C 672.

Sincerely,

Sandra Costello

Assistant Director, IRB # 2

Office for the Protection of Research Subjects

Enclosures:

1. **UIC Investigator Responsibilities, Protection of Human Research Subjects**
2. **Data Security Enclosure**

cc: Steven G. Rivkin, Economics, M/C 144

APPENDIX B. EVIDENCE OF COPYRIGHT

ELSEVIER LICENSE
TERMS AND CONDITIONS

May 03, 2016

This is a License Agreement between Irina Horoi ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Irina Horoi
Customer address	3055 N. Sheffield Apt. 3A
	CHICAGO, IL 60657
License number	3861551028249
License date	May 03, 2016
Licensed content publisher	Elsevier
Licensed content publication	Economics of Education Review
Licensed content title	Disruptive peers and the estimation of teacher value added
Licensed content author	Irina Horoi,Ben Ost
Licensed content date	December 2015
Licensed content volume number	49
Licensed content issue number	n/a
Number of pages	13
Start Page	180
End Page	192
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title of your thesis/dissertation	The Roles of Teacher and Peer Quality on Human Capital Accumulation in Primary and Secondary School
Expected completion date	May 2016
Estimated size (number of	120

pages)	
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the

materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:
15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.
16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following

clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.
17. **For journal authors:** the following clauses are applicable in addition to the above:
Preprints:
A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).
Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).
If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.
Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.
Authors can share their accepted author manuscript:

- immediately
  - via their non-commercial person homepage or blog
  - by updating a preprint in arXiv or RePEc with the accepted manuscript
  - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
  - directly by providing copies to their students or to research collaborators for their personal use
  - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- after the embargo period
  - via non-commercial hosting platforms such as their institutional repository
  - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.
Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.
If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.
Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.
Please refer to Elsevier's posting policy for further information.
18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.
19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

REFERENCES

Aaronson, D., L. Barrow, and W. Sander. "Teachers and Student Achievement in the Chicago Public High Schools" *Journal of Labor Economics*, 25, (2007): 95-135.

Aizer, Anna. "Peer Effects and Human Capital Accumulation: The Externalities of ADD." NBER Working Paper WP14354. Cambridge, MA: National Bureau of Economic Research, 2008.

Ballou, D., W. Sanders, and P. Wright. "Controlling for Student Background in Value-Added Assessments of Teachers." *Journal of Educational and Behavioral Statistics*, 29(1), (2004): 37–65.

Behrman, J. R., R. A. Pollak and P. Taubman. "Parental Preferences and Provision for Progeny." *The Journal of Political Economy*, 90(1), (1982): 52-73.

Cavalluzzo, Linda C. "Is National Board Certification An Effective Signal of Teacher Quality?" Institute of Education Sciences ED485515, 2004.

Cantrell, S., J. Fullerton, T.J. Kane, and D. O. Staiger. "National Board Certification and Teacher Effectiveness: Evidence from a Random Assignment Experiment." NBER Working Paper WP14608. Cambridge, MA: National Bureau of Economic Research, 2008.

Carrell, Scott and Mark L. Hoekstra. "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids," *American Economic Journal: Applied Economics*, 2(1), (2010): 211-228.

Chetty, R., J. R. Friedman and J. E. Rockoff[a]. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104(9), (2014): 2593-2632.

Chetty, R., J. R. Friedman and J. E. Rockoff[b]. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review*, 104(9), (2014): 2633-79.

Chetty, R., Friedman, J. N., Hilger, N. G., Saez, E., Schanzenbach, D. W., & Yagan, D. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star." *The Quarterly Journal of Economics, 126*(4), (2011): 1593-1660.

Chingos, M. M., & Peterson, P. E. "It's Easier to Pick a Good Teacher than to Train One: Familiar and New Results on the Correlates of Teacher Effectiveness." *Economics of Education Review, 30*(3), (2011): 449-465.

Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor. "Teacher-Student Matching and the Assessment of Teacher Effectiveness." *Journal of Human Resources, 41*(4), (2006): 778-820.

Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor. “Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects.” *Economics of Education Review, 26*(6), (2007): 673-682.

Cowan, James, and David Goldhaber. “National Board Certification and Teacher Effectiveness: Evidence from Washington.” Technical Report 2015-1, Center for Education Data and Research, Seattle, WA (2015).

Dee, Thomas, and James Wyckoff. "Incentives, Selection, and Teacher Performance: Evidence from IMPACT." NBER Working Paper WP19529. Cambridge, MA: National Bureau of Economic Research, 2013.

Doherty, Kathryn M. and Sandi Jacobs, The National Council on Teacher Quality.. State of States 2013 Connect the Dots: Using Evaluations of Teacher Effectiveness to Inform Policy and Practice (Research Report). (2013). Retrieved from http://www.nctq.org/dmsView/State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report

Feng, Li. “Hire Today, Gone Tomorrow: New Teacher Classroom Assignments and Teacher Mobility.” *Education Finance and Policy*, 5(3), (2010): 278-316.

Figlio, David N. “Boys Named Sue: Disruptive Children and their Peers.” *Education Finance and Policy,* 2(4), (2007): 376-394.

Fletcher, Jason M.[a] “The Effects of Inclusion on Classmates of Students with Special Needs: The Case of Serious Emotional Problems.” *Education Finance and Policy,* 4(3), (2009): 278-299.

Fletcher, Jason M.[b] “Spillover Effects of Inclusion of Classmates with Emotional Problems on Test Scores in Early Elementary School.” *Journal of Policy Analysis and Management,* 29(1), (2009): 69-83.

Friesen, Jane, Ross Hickey, and Brian Krauth. “Disabled Peers and Academic Achievement.” *Education Finance and Policy,* 5(3), (2010): 317-348.

Frijters, P., D. W. Johnston, M. Shah, and M. A. Shields. “Intrahousehold Resource Allocation: Do Parents Reduce or Reinforce Child Ability Gaps?” *Demography*, *50*(6), (2013): 2187-2208.

“Fiscal Year 2011-2012 North Carolina Public Schools Salary Schedule.” *North Carolina Public Schools Financial and Business Services.* North Carolina Department of Public Insruction. Web.

Goldhaber, Dan and Emily Anthony. “Can Teacher Quality be Effectively Assessed? National Board Certification as a Signal of Effective Teaching.” *The Review of Economics and Statistics*, 89(1), (2007):134-150

Harris, N. Douglas and Tim R. Sass. “The Effects of NBPTS-Certified Teachers on Student Achievement.” *Journal of Policy Analysis and Management, 28*(1), (2009): 55-80.

Harris, Douglas N. and Tim R. Sass. “Teacher Training, Teacher Quality and Student Achievement.” *Journal of Public Economics, 95*(7), (2011): 798-812.

Hanushek, Eric A. “The Evidence on Class Size” in Susan E. Meyer and Paul E. Petersib (ed.). *Earning and learning: How schools matter,* Washington, DC: Brookings Institution, 1999, 131-168.

Hanushek, Eric A. “Evidence, Politics, and the Class Size Debate” in Lawerence Mishel and Richard Rothstein (ed.). *The Class Size Debate,* Washington, DC: Economic Policy Institute, 2002, 37-65.

Hanushek, Eric A. “The Economic Value of Higher Teacher Quality.” *Economics of Education Review, 30*(3), (2011): 466-479.

Hanushek, Eric A. “Boosting Teacher Effectiveness” in Chester E. Finn Jr. and Richard Sousa(ed.). *What lies ahead for america’s children and their schools*, Stanford, CA: Hoover Institution Press, 2014, 23-35.

Hanushek, Eric A., J. F. Kain, and S. G. Rivkin. “Why Public Schools Loose Teachers.” *The Journal of Human Resources,* 39(2), (2004): 326-354.

Hanushek, Eric A., and Steve G. Rivkin. “Teacher Quality.” *Handbook of the Economics of Education, 2,* (2006):1051-1078.

Hakel, Milton D. "Teacher Participation in the Program." *Assessing Accomplished Teaching: Advanced-level Certification Programs: Committee on Evaluation of Teacher Certification by the National Board for Professional Teaching Standards*. Washington, DC: National Academies, 2008. 128. Print.

Horoi, Irina and Ben Ost. “Disruptive Peers and the Estimation of Teacher Value-Added.” *Economics of Education Review*, 49, (2015):180-192.

Ho, Andrew D., and Thomas J. Kane. “The Reliability of Classroom Observations by School Personnel.” Research Paper. MET Project. *Bill & Melinda Gates Foundation,* 2013.

Hoxby, Caroline M. “Peer Effects in the Classroom: Learning from Gender and Race Variation.” NBER Working Paper No. 7867, (2000).

Hoxby, Caroline M. and Gretchen Weingarth. “Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects.” Mimeo, Harvard University, (2005).

Imberman, Scott A., Adriana D. Kugler, and Bruce I. Sacerdote. "Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees." *American Economic Review*, 102(5), (2012): 2048-2082.

Isenberg, Eric and Elias Walsh. Measuring Teacher Value Added in DC, 2012-2013 School Year (Report No. 06038.502). Washington, DC: Mathematica Policy Research, (2013).

Jacob B. A., L. Lefgren and D. P. Sims. "The Persistence of Teacher-Induced Learning Gains." *Journal of Human Resources* 45(4), (2010): 915-943.

Johnson, M., S. Lipscomb, and B. Gill. "Sensitivity of Teacher Value-Added Estimated to Student and Peer Control Variables." Retrieved from Mathematica Policy Research Working Paper, (2013): http://vam.educ.msu.edu/wp-content/uploads/2013/10/Sensitivity-of-Teacher-VA-11_2013.pdf

Johnston, Howard. *Disruptive Behavior: School Based Interventions*. Retrieved from Education Partnerships Research into Practice, (2013): http://www.gearup.oous.edu/sites/default/files/research-briefs/researchbriefdisruptivebehavior.pdf

Kalogrides, D., Loeb, S., & Béteille, T. "Systematic Sorting Teacher Characteristics and Class Assignments." *Sociology of Education*, 86(2), (2013): 103-123.

Kane, Thomas J., "Do Value-Added Estimates Identify Causal Effects of Teachers and Schools?" The Brown Center Chalkboard, Brookings Institution, October 30, 2014.

Kane, T. J., & Staiger, D. O. "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation." NBER Working Paper WP14607. Cambridge, MA: National Bureau of Economic Research, 2008.

Kelly, Elaine. "The Scourge of Asian Flu in Utero Exposure to Pandemic Influenza and the Development of a Cohort of British Children." *Journal of Human Resources*, *46*(4), (2011): 669-694.

Koedel, Cory, and Jiaxi Li. "The Efficiency Implications of Using Proportional Evaluation to Shape the Teaching Workforce." *Contemporary Economic Policy*, 34(1), (2016): 47-62.

Kristoffersen, Jannie H. G., Morten V. Krægpøth, Helena S. Nielsen, and Marianne Simonsen. "Disruptive School Peers and Student Outcomes." *Economics of Education Review*, 45, (2015): 1-13.

Krueger, Alan B. "Economic Considerations and Class Size." *Economic Journal*, 113, (2003): 34-63.

Krueger, Alan B. and Diane M. Whitmore. “The Effect of Attending a Small Class in the Early Grades on College-test Taking and Middle School Test Results: Evidence from Project Star.” *The Economic Journal*, 111(468), (2001): 1-28.

Lavy, Victor, and Analia Schlosser. "Mechanisms and Impacts of Gender Peer Effects at School." *American Economic Journal: Applied Economics*, 3(2), (2011): 1-33.

Lavy, Victor, Daniele M. Paserman, and Analia Schlosser. “Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom.” *The Economic Journal,* 122(559), (2011): 208-237

Lazear, Edward. “Education Production,” *Quarterly Journal of Economics*, 116(3), (2001): 777-803.

Levy, Paul E. and Jane R. Williams. The Social Context of Performance Appraisal: A Review and Framework for the Future. *Journal of Management, 30*(6), (2004): 881-905.

Manski, Charles. “Identification of Endogenous Social Effects: The Reflection Problem.” *Review of Economic Studies*, 60(3), (1993): 531-542.

McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. “The Intertemporal Variability of Teacher Effect Estimates.” *Education Finance and Policy*, 4(4), (2009): 572-606.

McCaffrey, D.F., J. R. Lockwood, D. Koretz, T. A. Louis, and L. Hamilton. Models for Value-Added Modeling of Teacher Effects. *Journal of Educational and Behavioral Statistics* 29(1), (2004): 67-101.

Moffitt, Robert. “Policy Interventions, Low-Level Equilibrium, and Social Interactions,” in S. Durlauf and P. Young (Eds.), *Social Dynamics*. Cambridge, MA: MIT Press, 2001.

Neidell Matthew, and Jane Waldfogel. “Cognitive and Noncognitive Peer Effects in Early Education.” *The Review of Economics and Statistics,* 92(3), (2010): 562-576.

OECD. (2013). *How the quality of the learning environment is shaped.* Retrieved from http://www.oecd.org/pisa/keyfindings/Vol4Ch5.pdf

Ost, Ben and Jeffrey C. Schiman. “Grade-Specific Experience, Grade Reassignment, and Teacher Turnover.” *Economics of Education Review,* 46 (2015): 112-126.

Reardon, Sean F. “The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanation.” *Whither Opportunity?: Rising inequality, Schools, and Children’s Life Chances.* Ed. Greg J. Duncan and Richard Murnane. New York: Russell Sage Foundation, 2011, 91-116. Print.

Rivkin, S. G., E. A. Hanushek, & Kain, J. F. “Teachers, Schools, and Academic Achievement.” *Econometrica, 73*(2), (2005): 417-458.

Rosenzweig, Mark. R., and Paul T. Schultz. “Market Opportunities, Genetic Endowments, and Intrafamily Resource Distribution: Child Survival in Rural India.” *The American Economic Review*, 72(4), (1982): 803-815.

Rockoff, Jonah E. “The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data.” *American Economic Review*, 94(2), (2004): 247-252.

Rothstein, Jesse. Student Sorting and Bias in Value-Added Estimation: Selection on Observables and Unobservables.” *Education Finance and Policy,* 4(4), (2009): 537-571.

Rothstein, Jesse. “Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement.” *Quarterly Journal of Economics, 125*(1), (2010): 175-214.

Rothstein, Jesse (2014) “Revisiting the Impacts of Teachers.” Working paper. Retrieved from: http://eml.berkeley.edu/~jrothst/workingpapers/rothstein_cfr_oct2014.pdf

Royer, Heather. “Separated at Girth: US Twin Estimates of the Effects of Birth Weight.” *American Economic Journal: Applied Economics*, 1(1), (2009): 49-85.

Spence, Michael. “Job Market Signaling.” *The Quarterly Journal of Economics*, 87(3), (1973): 355-374.

Whitehurst, Grover J., Matthew M. Chingos and Katharine M. Lindquist. “Evaluating Teachers with Classroom Observations: Lessons Learned in Four Districts” Brown Center on Education Policy at Brookings, 2014.

Table I
The Effect of NBPTS Certification on Math and Reading Achievement
On the Full Student Sample[a,b]

	Math (1)	Reading (2)
Proportion NBPTS teachers in grade	0.038***	0.009**
	(0.005)	(0.003)
Observations	2845402	2821242
School-by-Year FE	Yes	Yes

Notes: [a]Estimates are presented from the preferred model on the full sample of students in years 2007-2013. All models include the student, teacher, and grade controls, which were specified in Table VII. Standard errors clustered at the school-by-grade-by-year are presented in parentheses (* p<0.05, ** p<0.01, *** p<0.001).
[b]Source Data: NCERDC

VITA

Name: Irina Horoi

Education:

Expected 2016	Ph.D. in Economics, University of Illinois at Chicago
2012	M.A., Economics, Central Michigan University
2008	B.A., Economics, University of Arizona

Fields:

Economics of Education, Labor Economics, Urban Economics, Applied Econometrics

Publications:

"Environmental correlates of physical activity and sedentary behavior in African American women: An ecological momentary assessment study" (with S. Zenk, K. Jones, C. Corte, B. Riley, J. Wilbur and L. Finnegan), *Women and Health,* 25, (2015): 1-17.

"Disruptive Peers and the Estimation of Teacher Value-Added" (with Ben Ost), *Economics of Education Review,* 49, (December 2015): 180-192

"Ecological momentary assessment of personal, environmental, and behavioral factors, and snack food intake in African American women" (with Shannon Zenk, Ashley McDonald, Colleen Corte, Barth Riley, and Angela Odoms-Young),
Appetite, 83, (December2014): 333-341

Working Papers:

"New Evidence of National Board Certification as a Signal of Teacher Quality" (with PhD candidate Moiz Bhai), Job Market Paper, 2015

"Teacher Characteristics and Student Achievement: Evidence from Twins" (with PhD candidate Moiz Bhai), 2015

Teaching Experience:

Spring 2015, 2016	Econometrics (undergraduate), UIC, Instructor
Fall 2014	Econometrics (undergraduate), UIC, Teaching Assistant
Spring 2013	Public Finance (graduate), UIC, Teaching Assistant
Fall 2012	Intermediate Macroeconomics (undergraduate), UIC, Teaching Assistant
Summer 2012	Principles of Microeconomics (undergraduate), UIC, Instructor
Spring 2012	Principles of Microeconomics (undergraduate), UIC, Instructor
Fall 2012	Principles of Microeconomics (undergraduate), UIC, Instructor

Spring 2011	Principles of Economics for Business (undergraduate), UIC, Teaching Assistant
Fall 2010	Principles of Microeconomics (undergraduate), UIC, Teaching Assistant

Research and Professional Experience:

2016	Amazon, Amazon Lending, Economist
2015	Edward Hines Jr. VA Hospital, Research Services, Research Assistant
2013-2015	UIC, College of Nursing, Research Assistant for Associate Professor Shannon Zenk
2014	UIC, Department of Economics, Research Assistant for Assistant Professor Javaeria Qureshi
2012-2014	UIC, Department of Economics Research Assistant for Assistant Professor Ben Ost
2012	UIC, Institute for Health Research Policy, Research Assistant for Professor Frank Chaloupka
Summer 2011	Mathematica Policy Research, Survey Operation Center

Academic Presentations

2015	The 143rd Meeting of the American Public Health Association (October), Economic Research Lunch Workshop at University of Illinois at Chicago (June), The Annual Conference of the Association for Education Finance and Policy (February), Active Living Research Conference (February)
2014	UIC Graduate Student Seminar (October), The Annual Conference of the Association for Education Finance and Policy (March)

Honors, Scholarships, and Fellowships:

2014	Gilbert Bassett, Barry Chiswick, Richard Kosobud, & Houston Stokes Award, UIC
2004-2007	Arizona Scholarship of Excellence
2004	Michigan Merit Award

Refereeing

Economics of Education Review

Professional Affiliations:

American Economic Association
American Public Health Association