New Evidence on National Board Certification as a Signal of Teacher Quality

Irina Horoi
Amazon

Moiz Bhai[*]
University of Arkansas at Little Rock

Abstract

Using longitudinal data from North Carolina that contains detailed identifiers, we estimate the effect of having a National Boards for Professional Teaching Standards (NBPTS) teacher on academic achievement. We identify the effects of an NBPTS teacher exploiting multiple sources of variation including the traditional lagged achievement models, twin and sibling effects, and aggregate grade level variation. Our preferred estimates show that students taught by National Board certified teachers have higher math and reading scores by 0.04 and 0.01 of a standard deviation. We find that an NBPTS math teacher increases the present value of students' lifetime income by $48,000.

JEL Codes: I20, I21, I22, I25

Keywords: student achievement, National Board certification, teacher quality, teacher effectiveness, twin and sibling designs

[*]University of Arkansas at Little Rock, Department of Economics, 2801 S. University Ave, Little Rock, AR 72204. Email: mxbhai@ualr.edu. We thank Pavel Dramski, Martha Bailey, Steven Durlauf, Ben Feigenberg, Nathan Hendren, Darren Lubotsky, Steven Rivkin, Ben Ost, Javaeria Qureshi, Jeff Smith and seminar participants at the University of Illinois at Chicago along with SSSI participants at the University of Chicago for helpful comments. We also thank the NCERDC for providing access and assistance to North Carolina data. All views expressed in this paper are our own, and the usual disclaimer applies.

1. 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, then 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. More 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. Such an approach strengthens the family design, because twins share more commonalities than 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 3^{rd}-8^{th} 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 0.012 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 disaggregated 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. It is not clear, however, 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.

2. Conceptual Framework

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 teacher's 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:

2.1 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, who 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

[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).

lifetime benefits are the same for both quality teachers, then a larger proportion of high quality teachers would enter the certification applicant pool.

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[3], 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.

2.2 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

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

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 achievement between certified teachers and those who never applied.[4] 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.[5]

2.3 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).

3. Identification Strategy

Studies that have examined the impact of National Board for Professional Teaching Standards (NBPTS) certification on student achievement, have generally found effects of

[4] While this study uses experimental data, it should be noted that the experiment was likely imperfect because the comparison group of non-certified teachers were chosen by the principal of the teacher's school and not randomly.
[5] 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.

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; Clotfelter 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. 2006, Clotfelter et al. 2007), student fixed effects (Harris and Sass, 2009), using schools where sorting is balanced on observables (Clotfelter, 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.[6] 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, like most studies conducted at the family level we are unable to test for parental responses. However, the literature on intra-family resource allocation is mixed, as some studies find that parents act by reinforcing differences (Frjjter et al., 2013; Rosenzwieg and Shultz, 1982), whereas others find that parents compensate for the inequality (Behrman 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,

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

within siblings-by-year, and twins-by-year.[7] 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 our preferred models using across-cohort comparisons within the same school. More 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 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-cohort 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.

---

[7] 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.

Thus, our study differs from earlier work on three notable dimensions. First, we employ alternative sources of identification using both within-family and across-cohort variation that might alleviate bias from the threat of non-random assignment of teachers to students. Second, the aforementioned studies that use North Carolina data rely on proctoring records to infer the teacher in the classroom, whereas we use data from the time when teachers are matched to students based on course membership records. Although random mismatches will only produce attenuation bias, matching students to proctors might introduce upward biases if a teachers' propensity to proctor is associated with teacher quality. The use of course membership records mitigates bias from both the former and the latter causes. Lastly, a notable difference is that we investigate the effectiveness of 2nd iteration NBPTS certification on student achievement as opposed to the original NBPTS standards.

4. Empirical Models

4.1 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.

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

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 expansion 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.[8] 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.[9] 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[10]. Similar to comparisons within twins, this specification accounts for unobservable differences between cohorts, in addition to accounting fixed and time varying differences between schools, neighborhoods, and families.

4.2 Grade Level – Cohort Variation

---

[8] 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.

[9] We also drop birth order and sibling spacing.

[10] We also exclude grade-by-year fixed effect as there is no variation in these dummies within a school-grade-year.

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[11,12] In addition, we estimate models where we include both 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.

5. **Data**[13]

[11] This is done for comparison purposes.
[12] To provide evidence that these estimates are generalizable, we run (2) on the full sample of students with just school-by-year fixed effects. The results are nearly identical to the same specifications on the sibling sample and can be found in appendix A, table 1.
[13] Some of the data description overlaps with Bhai and Horoi (2016).

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. We begin our time period in 2006 because teachers are directly matched to students based on course membership records instead of their proctors. The 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.

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 some differences 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.

6. Results

6.1 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.[14]

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

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

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

6.2 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. [15, 16] 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.[17] 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.

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

[15] Pre-observed student characteristics refer to characteristics observed with a one year lag.
[16] 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.
[17] Samples are constrained to school years with both certified and non-certified teachers. Column 3 is also constrained to only twins.

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.

6.3 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 our estimate 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 to differ from the general population of students in North Carolina.[18] 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.

6.4 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 with other non-certified teachers in their grade, and our preferred model does not parse out this

[18] 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.

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 who 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.

6.5 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.

6.6 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.[19] 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.

6.7 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

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

7. **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 explanations 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:[20]

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

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.

References

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.

Bhai, M., and Horoi, I. "What Makes for a Good Teacher?" Unpublished Working Paper, (2016).

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.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of human Resources*, *41*(4), 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.

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

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

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.

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).

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.

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

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.

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.

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.

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.

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.

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.

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.

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

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

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

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.

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.

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	**After**	
	Mean	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

[b]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			
Proportion of students in grade taught by NBPTS teachers	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

Appendix A

Table A.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