



National Board Certification and Teacher Effectiveness: Evidence from Washington

James Cowan and Dan Goldhaber

Center for Education Data & Research
University of Washington Bothell

February 11, 2015

Abstract: We study the effectiveness of teachers certified by the National Board for Professional Teaching Standards (NBPTS) in Washington State, which has one of the largest populations of National Board Certified Teachers (NBCTs) in the nation. Based on value-added models in math and reading, we find that NBPTS certified teachers are about 0.01-0.05 student standard deviations more effective than non-NBCTS with similar levels of experience. Certification effects vary by subject, grade level, and certification type, with greater effects for middle school math certificates. We find mixed evidence that teachers who pass the assessment are more effective than those who fail, but that the underlying NBPTS assessment score predicts student achievement. Finally, we use the individual assessment exercise scores to estimate optimal weights for value-added prediction.

Acknowledgements: This study was funded by the Bill and Melinda Gates Foundation and by the National Center for Analysis of Longitudinal Data in Education Research (CALDER), funded through grant #R305A060018 to the American Institutes for Research from the Institutes of Education Sciences, U.S. Department of Education. We thank both funders for their generous financial support. We thank Joe Doctor and the Washington Office of the Superintendent of Public Instruction (OSPI) for helpful comments and Christopher Tien for the expert research assistance he provided. We also thank OSPI and the National Board for Professional Teaching Standards for providing the data used in this study. Any and all errors are solely the responsibility of the study's authors, and the views expressed are those of the authors and should not be attributed to their institutions, the study's funders, or the agencies supplying data.

Suggested citation:

Cowan, J. and Goldhaber, D. (2015). National Board Certification: Evidence from Washington State. CEDR Working Paper 2015-3. University of Washington, Seattle, WA.

© 2015 by James Cowan and Dan Goldhaber. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission, provided that full credit, including © notice, is given to the source

You can access other CEDR publications at
<http://www.CEDR.us/publications.html>

Individual teachers have substantial influences on both immediate outcomes, such as standardized test scores, behavioral outcomes, and long-term outcomes, such as high school graduation, college attendance, and earnings.¹ Yet, the credentials typically rewarded in the labor market, advanced degrees and experience, do not explain much of the variation in teacher quality.² The National Board for Professional Teaching Standards (NBPTS), established in 1987, represents one strategy for recognizing teacher quality. The National Board is a voluntary system for assessing accomplished teaching. NBPTS offers an assessment process across several subject areas that is meant to signify teachers have achieved a high level of practice. NBPTS certification relies on an authentic, or “portfolio,” assessment process, which means that it uses artifacts of teacher practice, including videos of classroom lessons, student work, and reflective essays. Over the past two decades, both the program and the reach of National Board Certified Teachers (NBCTs) have grown substantially. Today, NBCTs number more than 100,000 and represent about 3 percent of the national teaching force (National Board of Professional Teaching Standards, 2010).

As of 2010, 30 states either offered financial incentives for teachers to complete the NBPTS assessment process or bonuses for certified teachers (Exstrom, 2011). Despite the extensive state interest in using the NBPTS assessment as a marker of teacher quality for human capital purposes, the extant research on the effectiveness of National Board Certified Teachers (NBCTs) has generated inconsistent results. Most of the studies using long longitudinal samples of students in states or districts with large populations of NBCTs have found that the difference in value-added between NBCTs and non-NBCTs of about 0.01-0.03 student standard deviations, which corresponds to about 20-30% of the returns to the first five years of teaching experience or about 2-10% of annual achievement gains in the elementary grades (Atteberry et al., 2013; Bloom et al., 2008; Harris and Sass, 2011; Wiswall, 2013).

We add to this literature with a study of NBCTs in Washington, a state with a large population of certified teachers that has not heretofore been studied. Our study is unique in that we consider heterogeneity in teacher effectiveness both by NBPTS assessment type and by whether candidates pass on their first attempt. We believe this is also one of only a few studies that use statewide data to specifically study the performance of teachers certified under the second generation NBPTS assessment regime introduced in 2002.³ We find that teachers who possess the National Board credential are about 0.01-0.05 standard deviations more effective than non-NBCTS with similar levels

¹ See Aaronson et al. (2007), Chetty et al. (2014a, 2014b), Jackson (2012), Nye et al. (2004), Rivkin et al. (2005).

² See Goldhaber et al. (1999), Goldhaber and Hansen (2013), Harris and Sass (2011), and Kane et al. (2008).

³ Harris and Sass (2009), who break out NBCTs by their licensure cohort and include some cohorts licensed under both the first and second generation of assessments, find some evidence of differential effects by cohort. Chingos and Peterson (2011) study teacher credentials in Florida between 2002 and 2009, but do not explicitly break out NBPTS credentials by certification type.

of experience depending on the classroom level and subject. Comparing our results to the average achievement gains estimated from vertically aligned, nationally normed assessments, we estimate that NBCTs produce annual learning gains that are about 4-5% of normal learning gains at the elementary school level, about 15% of annual learning gains in middle school math, and about 4% of annual learning gains in middle school reading (Bloom et al., 2008). We additionally find evidence that performance on the most common certificates at the elementary and middle school levels predicts student achievement. Finally, the National Board for Professional Teaching Standards allows candidates who initially fail the assessment to bank their scores and retake portions of the examination process. In our data, teachers who initially failed represent about 30% of NBCTs. We therefore consider the effectiveness of National Board candidates based on whether they gained certification on their first attempt or on a retake. Except in middle school mathematics, we do not find evidence that teachers earning certification through a retake are more effective than non-NBCTs.

I. Background and Previous Findings on NBPTS Teachers

The National Board for Professional Teaching Standards was established in 1987 to offer a national teaching credential signifying the accomplishment of a high level of professional teaching. As National Board Certification is one of the few *national* teaching credentials in the United States, prior research has documented the effectiveness of NBCTs in several states.⁴ The relatively small body of literature on average differences in value-added by NBCT status has thus far yielded mixed results using states or districts with large populations of NBCTs. On the other hand, the few papers that have assessed differences in teacher effectiveness within the pool of NBCT applicants have found clearer evidence that teachers who do better on the NBPTS assessment tend to be more effective teachers.

Observational studies of NBCT effects have generally yielded point estimates in the range of 0.01-0.03 standard deviations on statewide assessments, or about 2-10% of an average year's learning gains, with not all studies finding statistically significant effects. In a study of elementary classrooms in North Carolina, Goldhaber and Anthony (2007) find that NBCTs raise student achievement in reading by about 0.02 standard deviations more than non-NBCTs with similar credentials. Results for math are smaller and statistically insignificant.⁵ They additionally find that recently certified NBCTs appear to be about 0.06-0.08 student standard deviations more effective with poor children, although this result does not appear to hold for teachers certified in previous years. Using a longer panel of elementary school data from North Carolina, Clotfelter et al. (2007) estimate statistically significant effects of 0.02-0.03 standard deviations for

⁴ As of 2010, 39 states accept the NBPTS credential as a means to fulfill state licensing or continuing education requirements (Exstrom, 2011).

⁵ On the other hand, they consistently find that future NBCTs are more effective than teachers who never become certified.

certified teachers in math. In reading, the effects are about 0.01 standard deviations, but the statistical significance varies by the model specification. However, in a companion paper that focuses more intently on the potentially non-random sorting of students to teachers in elementary school classrooms, Clotfelter et al. (2006) find no evidence of NBCT effects in their most conservative models. Among high school teachers in North Carolina, Clotfelter et al. (2010) find that NBCTs are about 0.05 standard deviations more effective than non-certified teachers. Evidence from Florida, another state with a large NBCT population, is also mixed. Chingos and Peterson (2011) document positive effects of NBCTs of about 0.02-0.03 standard deviations in both math and reading on the FCAT. Harris and Sass (2009) find no general effect of NBCTs, but do find some statistically significant results depending on the certification cohort and test. In the only existing experimental evaluation of NBCT effectiveness, Cantrell et al. (2008) find no statistically significant differences between students in classrooms randomized to NBCTs and those in classrooms randomized to non-applicants. However, compared to the statewide longitudinal samples in other research, their randomized sample contains a relatively small number of certified teachers.

The NBCT effects estimated in the above papers compare successful applicants for board certification both to unsuccessful applicants *and* to teachers who never apply for certification. If teachers who apply for certification are more effective than other teachers, the observed NBCT effects may be due to the selection of teachers who apply for certification rather than to the discrimination of the actual assessment process. Alternatively, if less effective teachers tend to apply, the above findings would understate the power of the NBPTS process to discern differences in teachers' value added. While the results comparing certified and non-NBCTs are mixed, it appears that the NBPTS assessment does differentiate between more and less effective teachers. Goldhaber and Anthony (2007) find that successful applicants are about 0.13 standard deviations more effective in math and about 0.07 standard deviations more effective in reading than unsuccessful applicants. And Cantrell et al. (2008) find that successful applicants outperform unsuccessful applicants by about 0.22 standard deviations in math and 0.19 standard deviations in reading. They further find that the scaled score predicts student achievement in both subjects, with a one standard deviation difference in performance on the NBPTS assessment translating into a 0.11 standard deviations difference in student achievement in math and a 0.05 standard deviations difference in reading.

In sum, point estimates suggest that NBCTs are about 0.01-0.03 standard deviations more effective than non-NBCT elementary school teachers, with mixed statistical significance. An effect of this size is comparable to roughly 20-30% of the returns to the first five years of teaching experience or about 2-10% of annual student achievement gains in reading (Atteberry et al., 2013; Bloom et al., 2008). While the difference in value-added between NBCTs and non-NBCTs may vary by state, subject, and grade level, it does appear that performance on the assessment predicts student achievement.

II. Data

We base our study of National Board teachers on data from Washington State. Although Washington has only the 15th largest population of K-12 public school students in the United States, it has the fourth most NBCTs of any state and produced the most newly certified teachers in 2014 (National Board of Professional Teaching Standards, 2014a, 2014b; Snyder and Dillow, 2013). This is likely due in part to the fact that Washington incentivizes National Board certification in a number of ways. In 2000, the state introduced a bonus of 15% of base salary for NBCTs.⁶ This was changed to \$3,500 in 2002 and \$5,000 in 2008. In the same year, the state introduced the Challenging Schools Bonus, an additional \$5,000 bonus for NBCTs working in high-poverty schools.⁷ Both the state and districts provide various incentives and support for NBPTS candidates. The state also provides a \$2,000 conditional loan for teachers who apply for certification, awards professional development credit for participation, and considers National Board Certification an acceptable way to satisfy the state's advanced certification requirement.⁸ Many districts offer their candidates additional incentives in the form of financial support, release for certification activities, or mentoring. Since the introduction of the bonuses, the number of NBCTs has increased dramatically. Between 2008 and 2012, the cumulative number of NBCTs statewide increased from 2,703 to 6,739 (National Board of Professional Teaching Standards, 2012).

We obtain teacher records in Washington State from the S-275, which is a survey of district personnel by the Office of the Superintendent of Public Instruction (OSPI). The S-275 contains information on teacher demographic characteristics, such as age, sex, and ethnicity, and teacher credentials, such as experience and educational attainment. Pearson, which manages the assessment of teacher candidates for NBPTS, provided us with a database of assessment results for teachers in Washington State. We matched the NBPTS data to the S-275 using full name and date of birth. We matched 94% of NBCT candidates working in public schools using full name and date of birth and an additional 4% using last or maiden name, first initial, and date of birth. Minor misspellings of names in the S-275 data are not uncommon; we additionally matched by hand another 1% of candidates using names, dates of birth, and schools of employment. Overall, we matched 12,189 of the 12,309 NBPTS candidates (99%) to employment records in the S-275.

In this study, we analyze candidates for all of the certificates offered by the NBPTS. However, we focus much of the analysis on four of the most common certificates at the elementary and middle school levels: the Middle Childhood: Generalist (MC/Gen), Early/Middle Childhood: Literacy, Reading and Language Arts (EMC/LRLA), Early

⁶ Throughout this paper, we refer to school years by the calendar year of the spring term.

⁷ The Challenging Schools Bonus pays teachers a maximum of \$5,000 and is prorated by the amount of time a teacher spends in an eligible school.

⁸ Washington revised its certification process in 2000 and accepts the National Board certificate as a substitute for the requirements for the "Professional" teaching certificate, which requires teachers to complete a portfolio assessment.

Adolescence: English Language Arts (EA/ELA), and Early Adolescence: Math (EA/Math) certificates. These account for 43% of the certificates awarded in Washington State. Because the NBPTS assessment process changed in the early 2000s, we additionally focus on teachers certified under the second-generation assessment process, which account for most of the NBCTs in Washington.⁹

We obtain student records from student longitudinal databases maintained by OSPI. The state requires standardized testing in math and reading in grades 3-8, and these test scores form the basis of our analysis. For school years 2006 to 2009, the student data system included information on students' registration and program participation but did not explicitly link students to their teachers. We therefore matched these students to teachers using the proctor identified on the end-of-year assessment. To ensure that these are likely to represent students' actual teachers, we limit the 2006-2009 sample to elementary school classrooms (grades 4-6), which tend to be self-contained, with between 10 and 33 students where the identified teacher is listed in the S-275 as 0.5 FTE in that school, taught students in no more than one grade, and is endorsed to teach elementary education.¹⁰ Between 2009-2010 and 2012-2013, the student longitudinal data system explicitly links students to their teachers in all grades. Our sample therefore additionally includes classrooms in grades 6-8 for these school years.¹¹

We present summary statistics for our analytical dataset in **Table 1**. Despite the large incentive to teach in high-poverty schools, at both the elementary and middle school level, National Board Certified teachers have classrooms with significantly higher baseline student achievement. In elementary grades, students of NBCTs have baseline achievement of about 0.05 standard deviations higher in math and 0.03 standard deviations in reading than those of non-NBCTS. At the middle school level, students of NBCTs have baseline achievement 0.17 standard deviations higher in math and 0.10 standard deviations higher in reading. The demographic composition of classrooms taught by NBCTs and non-NBCTs is similar.

⁹ That is, when we break out certificates by type, we only consider teachers certified under the second generation assessment who received certificates between 2002 and 2013. Therefore, some teachers with "other" certificates possess an earlier version of the same certificate. Given the small number of teachers certified in Washington before 2002, this does not encompass many teachers.

¹⁰ Some of the data related to students and teachers used in this study are linked using the statewide assessment's "teacher of record assignment", a.k.a. assessment proctor, for each student to derive the student's "teacher". The assessment proctor is not intended to and does not necessarily identify the sole teacher or the teacher of all subject areas for a student. The "proctor name" might be another classroom teacher, teacher specialist, or administrator. For the 2009-2010 school year, we are able to check the accuracy of these proctor matches using the state's new Comprehensive Education Data and Research System (CEDARS) that matches students to teachers through a unique course ID. Using the restrictions described above, our proctor match agrees with the student's teacher in the CEDARS system for about 95% of students in both math and reading.

¹¹ As some schools in Washington State use self-contained classrooms in grade 6, we split the sample based on the class type rather than the grade level. Both elementary and middle school samples therefore include some students in 6th grade.

At the elementary level, the MC/Generalist certificate is by far the most common. In our sample, 7 percent of all classrooms and 71 percent of classrooms taught by an NBCT are taught by a teacher holding this credential. Also common is the EMC/LRLA certificate, which accounts for 18 percent of all classrooms taught by an NBCT. For middle school students, the EA/Math and EA/ELA certificates are the most common. Among all math classrooms, 9 percent are taught by an NBCT, and 7 percent are taught by a teacher with the EA/Math credential. In reading, NBCTs teach 11 percent of middle school classrooms, and teachers with an EA/ELA certificate teach nearly 7 percent of classrooms.

III. Board Certification and Teacher Effectiveness

Following prior research on the student achievement effects of teacher characteristics, we estimate a value-added model that includes teachers' National Board certification status:

$$A_{ijt} = \rho A_{ijt-1} + X_{ijt}\beta + NBCT_{jt}\delta + T_{jt}\gamma + X_{jt}\pi + \epsilon_{ijt} \quad (1)$$

We control for lagged achievement using a vector that includes a cubic expansion of prior test scores in both math and reading. We additionally include in X_{ijt} student gender, race and ethnicity, FRL eligibility, learning disabled status, and participation in special education, English language learning, or gifted programs; we include in X_{jt} the teacher-year means of all of these variables.¹² In our most basic model, $NBCT_{jt}$ simply indicates whether teacher j is an NBCT in year j . In some models, we replace the NBCT indicator with a vector indicating the teachers' certificate area. The vector T_{jt} includes an indicator for each year of experience. In all models, we cluster standard errors at the teacher level. As the NBPTS assessment relies on artifacts of student learning from a teacher's classroom, we drop all school years in which teachers submitted an NBPTS portfolio in order to avoid mechanical correlation between the assessment results and student achievement. We additionally estimate models with both school and school-by-grade-by-year (cohort) fixed effects in order to explicitly make comparisons of NBCTs to other teachers in the same school. The state incentive program for NBCTs to work in high-poverty schools may bias estimates of the NBCT effect if attendance at such schools is associated with unobserved factors that influence student achievement.

Consistent estimation of the NBCT effect in Eq. (1) requires student assignment to an NBCT to be exogenous conditional on the student characteristics included in X . Whether teacher assignments satisfy this assumption in practice remains a contentious

¹² Using district-level data that permits better identification of discrete classrooms, Johnson et al. (2014) find that teacher value-added models that rely on teacher-year means of control variables produce teacher effects estimates with correlations of between 0.93 and 0.98 with models using classroom means.

point. At the elementary level, Rothstein (2010) presents evidence of sorting into future classrooms based on unobserved shocks to student achievement. However, such empirical findings may be consistent with assignment policies that result in relatively unbiased estimates of teacher effects, and there is some experimental and quasi-experimental evidence that this is the case (Chetty et al., 2014a; Goldhaber and Chaplin, 2014; Kane et al., 2013; Kane and Staiger, 2008). However, grouping of students by ability may be more common at higher grade levels, and such tracking may still bias estimates of teacher effects (Jackson, 2014; Protik et al., 2013).

Even if value-added measures produce unbiased predictions of future student achievement on average, it remains possible that teacher effects are biased for certain subgroups of teachers. There are two related threats to validity in the context of estimating NBCT effects. First, as shown in Table 1, NBCTs teach students with higher lagged achievement, particularly at the middle school level. To the extent that measured student performance is correlated with unobserved contemporaneous inputs, estimated NBCT effects may be biased upward. For instance, higher-achieving students assigned to NBCTs may have greater intrinsic motivation or may receive better extracurricular or home instruction. Second, NBCTs are also more likely to teach gifted and honors students and, at the middle school, less likely to teach special education students. Even if such students do not differ in unobservable ways from similar students not assigned to such courses, there may still be effects associated with the grouping of such students in classrooms. These may be due to specific interventions, like assignment to better teachers in other subjects or access to additional school resources, or due solely to the exposure to higher-achieving peers (Jackson, 2014; Lavy et al., 2012; Lefgren, 2004). While some of the grouping effects may be captured by including teacher-year averages of lagged achievement measures, the classroom peer effects may not be constant across the student ability distribution. For instance, higher achieving students may benefit disproportionately from enrolling in classes with other high achieving students (Burke and Sass, 2013; Duflo et al., 2011). Thus, inclusion of peer characteristics alone may fail to capture important unobserved differences across classroom types that are associated with teacher certification status.

We implement two approaches aimed at generating comparisons of NBCTs to other teachers who teach in similar classrooms. For the elementary school classrooms, we follow the approach of Clotfelter et al. (2006) and re-estimate our models with cohort effects on samples of schools for which there is little evidence of classroom sorting by observable student characteristics, i.e. the demographic breakdown of classrooms in a school looks similar to the student demographics of the whole school. We classify students according to their prior test scores, gender, race, ethnicity, and participation in gifted, ELL, or special education programs and conduct chi-square tests assuming equal representation of students across classrooms within the same school, grade, and year.¹³ In

¹³ Our chi-square tests include indicators for whether the student scored above the median on each of the

our analysis sample, we use cohorts for which we have at least two classrooms and fail to reject all eight hypothesis tests as our restricted sample.¹⁴ As classrooms at the middle school level are much more likely to exhibit evidence of sorting on observables, this approach becomes untenable. Instead, to account for the possibility that student grouping or track-based interventions bias our estimates of the NBCT effects, we follow the approaches of Jackson (2014) and Protik (2013) and include cohort-by-track fixed effects for our middle school sample.¹⁵ This approach limits comparisons of NBCTs to other teachers in the same school, grade, and year who also teach students of the same level. Thus, we assume that omitted peer effects or track-based interventions have constant effects across classrooms within tracks and cohorts.

We present the results of these models for elementary classrooms in **Table 2**. In models with controls for observed student and classroom covariates, we find that NBCTs are 0.035 standard deviations more effective in math and 0.027 standard deviations more effective in reading than the average teacher with similar experience. In our preferred specification, which includes school-by-grade-by year fixed effects, these coefficients decrease to about 0.02 standard deviations for both math and reading.¹⁶ For the sample with balanced classrooms and cohort fixed effects, the coefficients are similar to those in the full sample, albeit not statistically significant in reading. We estimate coefficients of 0.018 in math, which is only statistically significant at the 0.10 level, and 0.007 in reading. The majority of the NBCTs in our elementary school sample (70%) have the Middle Childhood: Generalist (MC/Gen) certificate. We find that these teachers are 0.02 standard deviations more effective in math and 0.01 standard deviations more effective in reading than the average teacher; however, only the math result is statistically significant at the 0.05 level. Nearly 20% of certified teachers hold the Early and Middle Childhood: Literacy, Reading, and Language Arts (EMC/LRLA) certificate. We estimate an effect of

state standardized tests from the prior year; whether the student is female; whether the student is white; whether the student participates in gifted programs; whether the student participates in ELL programs; and whether the student participates in special education programs.

¹⁴ Clotfelter et al. (2006) pool estimates to the school level using classrooms in grades 3-5 in one school year. As they point out, the chi-square test may lack power to detect if schools do in fact sort students. To test whether we are actually identifying cohorts with balanced classrooms, we regress the baseline student characteristics on cohort and classroom fixed effects in the restricted sample and test the joint significance of the classroom fixed effects. Using a p-value of 0.10 in the chi-square tests to determine non-random assignment, we find that none of the models rejects the null hypothesis of no classroom effects at any conventional level.

¹⁵ Jackson (2014) uses a finer designation of tracks at the high school level by using groups of students who take the same courses. As our dataset does not permit the identification of individual courses at the middle school level, we follow Protik et al. (2013) and use indicators for course type to identify tracks. In our data, we identify a track as a unique combination of school, grade, school year, honors status, and remedial status. Honors and remedial courses are not identified at the elementary school level.

¹⁶ Because they implicitly limit comparisons of NBCTs to teachers within the same school and grade, models with cohort effects may be conservative estimates if there are differences in true teacher effectiveness across schools.

certified teachers of about 0.025 standard deviations in both subjects with the reading result statistically significant.

The results of the middle school analysis are described in **Table 3**. The middle school math results suggest that middle school NBCTs are somewhat more effective than average teachers and have a greater effect than elementary school NBCTs. We find that NBCTs are about 0.05 standard deviations more effective in teaching middle school math than non-certified teachers with similar levels of experience. Both results are robust to the inclusion of cohort and track fixed effects. When we disaggregate by certificate type, we find the coefficient on Early Adolescence: Math (EA/Math) drives the larger effect in the middle school math sample. These teachers comprise about 70% of our board certified teachers and are, on average, 0.065 standard deviations more effective than non-certified teachers. Overall, NBCTs are 0.01 standard deviations more effective than the average teacher in middle school reading education. The most common certificate at this level is the Early Adolescence: English Language Arts (EA/ELA) certificate (62%), and teachers who possess this credential are about 0.013 student standard deviations more effective than non-NBCTs.¹⁷

The NBPTS allows candidates who fail their assessment to bank their scores and reattempt one or more exercises. Because candidates can keep the scores from exercises in which they did particularly well and drop the exercises in which they did particularly poorly, it may be easier to earn certification on a retake than if candidates were forced to resubmit an entirely new application. We explore whether candidates who initially fail the assessment but later earn certification are more effective than non-NBCTs in Panel C in both Tables 2 and 3. We replace the indicator for NBCTs with an indicator for a teacher who has earned certification on the first attempt and an indicator for a teacher who has earned certification on a subsequent attempt.¹⁸ These models therefore compare NBCTs who earn certification on a first attempt and those who earn certification on a subsequent attempt to teachers who never earn certification. For elementary classrooms and middle school reading, we find two sets of common findings. First, we do not find evidence that initially unsuccessful applicants that go on to earn certification are more effective than non-NBCTs. The coefficients are small or negative and not statistically significant. Second, it appears that NBCTs who were initially unsuccessful applicants are less effective than NBCTs who earn certification on their first attempt. Tests of equality of the

¹⁷ An open question is whether participation in the National Board process improves teacher practice. We additionally estimate models that include teacher fixed effects and a censored experience profile at 10 years to test whether participation in the National Board process improves teacher value-added. We find small and imprecisely estimated within-teacher differences in effectiveness. These results are consistent with most of the prior results using student test score data and specifications with teacher fixed effects (Chingos and Peterson, 2011; Goldhaber and Anthony, 2007; Harris and Sass, 2009). Results are available from the authors upon request.

¹⁸ At the elementary school level, 4.9% of students have an NBCT who earned certification on the first attempt and 1.7% have an NBCT who earned certification on a retake. At the middle school, level these numbers are 8.1% and 2.6% for math and 9.9% and 3.0% for reading.

coefficients on passing on the first and subsequent attempt reject the hypothesis that the two groups are equally effective at the 10% level for all three subject-grade level groups.¹⁹ However, these results do not hold for middle school math teachers: those who pass the NBPTS assessment on a second take are still about 0.04 standard deviations more effective than other middle school math teachers. Furthermore, we fail to reject the hypothesis that the two groups of NBCTs are equally effective.²⁰ While there is some variation by certificate type, it appears that the first attempt generally contains more useful information about teacher effectiveness than subsequent attempts, which is consistent with Cantrell et al. (2008). We revisit this question in the section on NBPTS assessment results below.

Overall, we find that certified teachers are more effective than non-certified teachers with similar experience. The differences in average value-added range from 0.01-0.05 standard deviations depending on the subject and level. Our estimates for elementary school teachers in math and reading are of the same magnitude as those found for teachers in North Carolina (Clotfelter et al., 2007; Goldhaber and Anthony, 2007) and Florida (Chingos and Peterson, 2011). For middle school teachers, our results for the EA/Math certificate are closer in magnitude to those found at the high school level (Clotfelter et al., 2010), while the effects for teachers credentialed under the EA/ELA assessment are similar to the results for elementary school teachers. The additional learning gains produced by NBCTs for elementary students and middle school reading students are approximately 3-5% of annual achievement growth, while those produced by NBCTs in middle school math represent about 15% of annual learning gains in math (Bloom et al., 2008). This suggests NBCTs produce additional learning gains of about 1-2 weeks at the elementary school level and for middle school reading and about 5 weeks for middle school math.²¹

Exploring Heterogeneity in NBPTS Effects Across Student Sub-Groups

The National Board standards include the proposition that teachers should understand how to assess student learning and employ instructional techniques appropriate for their particular students. Teachers certified by the National Board may therefore be particularly adept at teaching students with extraordinary needs. Prior research suggests that National Board teachers are more effective with disadvantaged students and that participation in the National Board certification process improves

¹⁹ Note that these are two-sided tests. For models with cohort fixed effects, the F-statistic for the test of the equality of the coefficients is $F = 11.6$ ($p < 0.01$) for elementary math, $F = 3.75$ ($p = 0.05$) for elementary reading, and $F = 3.48$ ($p = 0.06$) for middle school reading. When we stack data across elementary and middle schools, we reject the hypothesis that the two groups are equally effective at the 5% level in both math and reading.

²⁰ The F-statistic from the test of equality of the coefficients is $F = 1.67$ ($p = 0.20$) for middle school math.

²¹ We convert gains on standardized tests to weeks or months of learning by averaging the results of Bloom et al. (2008) over the relevant grade range and assuming a 36 week school year. These results suggest annual learning gains of 0.50 and 0.36 standard deviations for elementary math and reading, respectively, and 0.34 and 0.27 standard deviations for middle school math and reading.

teachers' student assessment skills (Goldhaber and Anthony, 2007; Sato, Wei, and Darling-Hammond, 2008).

The relative efficacy of NBCTs for disadvantaged student subgroups has particular policy relevance. Previous work has documented that schools with large populations of impoverished children tend to have fewer NBCTs (Goldhaber, 2006; Humphrey et al., 2005). This finding is consistent with other evidence, based both on observed teacher credentials and teacher value-added, that high-quality teachers are not equitably distributed across or within schools (Clotfelter et al., 2007; Chetty et al., 2014a; Goldhaber et al., 2014; Sass et al., 2012). Yet, Koppich et al. (2007) suggest that teacher quality in low-performing schools was an early concern of the NBPTS and that some of its founders believed states or districts might develop financial incentives for NBCTs to teach in high-needs schools. In Washington State, NBCTs have been awarded a \$5,000 bonus since 2008 to teach full-time in high-poverty schools. Such policies at least implicitly assume that the effectiveness of NBCTs observed generally carry over to students in high-poverty schools.

In order to better understand the effectiveness of NBCTs for disadvantaged students, we add interactions between student characteristics and the NBCT indicator in Eq. (1):

$$A_{ijt} = \rho A_{ijt-1} + X_{ijt}\beta + NBCT_{jt}\delta + NBCT_{jt} \times X_{ijt} \lambda_{subgroup} + T_{jt}\gamma + \epsilon_{ijt}. \quad (2)$$

In Eq. (2), the $\lambda_{subgroup}$ test whether National Board Certified teachers are more or less effective for particular groups of students. We include interactions between NBCTs and indicators for gifted and talented students, English language learners, students receiving special education services, and students eligible for free and reduced price lunches. As with Eq. (1), the regression models additionally include school-by-year-by-grade effects.

The interaction effects $\lambda_{subgroup}$ estimated in Eq. (2) give the average difference in achievement for students of the given subgroup relative to other students with an NBCT. The total effect of NBCTs for that subgroup can be obtained by summing $\delta + \lambda_{subgroup}$. Thus, a negative coefficient $\lambda_{subgroup}$ suggests that students of the particular subgroup have lower achievement than other students assigned to an NBCT; only if the aggregate effect $\delta + \lambda_{subgroup}$ is negative would we conclude that students of this subgroup have lower achievement than other students of the same subgroup assigned to a non-NBCT. Supposing our estimates reflect the causal contributions of teachers to student learning, there are two possible explanations for finding evidence of differential effects of NBCTs for certain student subgroups. First, it may be the case that the teaching skills assessed by the NBPTS process are differentially important for students with particular needs. For instance, Sato et al. (2008) suggest that the certification process improves teachers' ability to use student assessment to support instruction. Alternatively, it may be the case that the most effective NBCTs are more likely to be assigned to certain kinds of students. Suppose

we find a positive interaction between NBCT status and giftedness. It may not be the case that individual NBCTs are more effective for gifted students, but that the more effective NBCTs are more often assigned to teach gifted students. This second possibility is consistent with the evidence on the within-school variation in teacher quality (Goldhaber et al., 2014). In order to differentiate between these two possibilities, we additionally estimate Eq. (2) with classroom fixed effects to control for any fixed teacher quality component.²² The interaction terms $\lambda_{\text{subgroup}}$ in these models compare the differences in achievement between students of particular subgroups and the reference category in NBCT classrooms to the difference in achievement between students of particular subgroups and the reference category in non-NBCT classrooms.²³ For instance, if the difference in achievement (conditional on prior test scores and other covariates) between gifted and non-gifted students is larger in NBCT classrooms than non-NBCT classrooms, we would conclude that NBCTs are relatively more effective at teaching gifted students.

We present the results of the student-level heterogeneity regressions in **Table 4**. In general, we find mixed evidence regarding student disadvantage and NBCT effectiveness. In elementary classrooms, we find that NBCTs are about 0.03 standard deviations less effective with English language learners than with other students in reading (the point estimate is negative but statistically insignificant in math). The estimated interaction is nearly identical when we include classroom fixed effects, which suggests that this reflects something about the teaching methods employed by NBCTs rather than their classroom assignments. With the inclusion of classroom fixed effects, we also find that elementary teachers are about 0.05 standard deviations more effective with gifted students, about 0.02 standard deviations more effective with special education students, and about 0.02 standard deviations less effective with FRL students than with omitted students. The same patterns are reflected in the reading results, although the coefficients are not statistically significant. Because the state incentive policy likely affects the distribution of NBCTs across student demographic groups and this may influence our findings, we also estimate models in columns (3) and (6) that additionally include an interaction between the NBCT effect and an indicator for school-wide eligibility for the Challenging Schools Bonus. The interaction with the Challenging Schools Bonus is positive but not statistically significant, which suggests that the difference in teacher effectiveness between NBCTs and non-NBCTs in challenging schools is similar to other schools.

At the middle school level, there is less evidence of subgroup heterogeneity. We find that NBCTs are actually less effective with special education students in math at the middle school level by about 0.04 student standard deviations, and this difference persists when we include classroom fixed effects. The discrepancy between the elementary and middle school math results may reflect differences in the curriculum or the fact that only

²² Specifically, we control for teacher-by-track fixed effects, which may not uniquely identify classrooms in middle schools (Johnson et al., 2014).

²³ In this case, the reference category is students not receiving gifted, English language learner, special education, or FRL services.

about half as many students are labeled as receiving special education services at the middle school level. As with the elementary school case, when we rely on within-classroom variation, we find that NBCTs are about 0.02 standard deviations less effective with FRL students. For middle school reading, none of the interaction terms is statistically significant. As with the results for elementary school teachers, we find that NBCTs at challenging schools are no more effective than NBCTs in other schools.

Our estimates of the subgroup heterogeneity of NBCT effects are somewhat at odds with prior research on the subject. Goldhaber and Anthony (2007) find that NBCTs appear to be more effective with FRL-eligible students, while Harris and Sass (2009) find little evidence, positive or negative, of effects for FRL students. By contrast, our estimates suggest that, in mathematics, NBCTs produce smaller gains for FRL students and, in reading, for English language learners, than other students in the same classroom. These results are robust to the sample of schools with apparently random assignment at the elementary level. While it is unclear what drives the differences in subgroup effects we estimate, there are two important policy differences between the North Carolina and Washington studies worth mentioning. First, Goldhaber and Anthony (2007) study teachers certified under the first-generation NBPTS assessment, which placed less emphasis on the assessment center exercises. Second, Washington incentivizes NBCTs to work in low-income classrooms, which may affect the distribution of NBCTs across classrooms.

IV. National Board Assessment Results and Teacher Effectiveness

Student Achievement Along the NBPTS Assessment Distribution

Although policymakers may be interested in the signaling value of the National Board certificate, the credential effects we estimate above may not accurately represent how well the assessment process discriminates between effective and ineffective candidates because the sample of NBPTS candidates is not randomly selected from the population of teachers. Therefore, we also assess the relationship between teacher value-added and the NBPTS assessment results. There are two potential complications with the estimation of the association between teacher value-added and performance on the assessment. First, the National Board assessment relies on evidence from student work and places particular emphasis on how teachers assess their students' progress (Pearlman, 2008). The portfolio design therefore introduces a possibly spurious correlation between measured teacher value-added and student achievement if raters' assessments of teacher practice are influenced by the students selected for inclusion in the NBPTS portfolio. As with the results on certified teachers, we therefore estimate models that exclude classrooms with a teacher who is participating in the National Board assessment process.

A second concern is that teacher performance may vary over time. While most research on the returns to teacher experience document substantial increases in teacher effectiveness during the first few years in the classroom, the returns to experience are

much smaller over the portion of the career in which teachers obtain certification (Papay and Kraft, 2013; Rockoff, 2004). However, recent research also suggests that long-run teacher effects are not perfectly persistent across time (Chetty et al., 2014a; Goldhaber and Hansen, 2013). We may therefore expect that the correlation between NBPTS assessment results and teacher value-added measured in different years understates the true contemporaneous correlation. In order to account for this possibility, we restrict our analysis of assessment results to years near participation in the National Board assessment process. In particular, we use classrooms for which the teacher completes a submission in years $t-2$, $t-1$, $t+1$, or $t+2$.²⁴

We begin by estimating the difference in value-added between teachers who initially pass and fail the National Board assessment. Using data on the classrooms of teachers who apply for certification, we regress achievement on student characteristics and an indicator for passing the National Board assessment:

$$A_{ijt} = \rho A_{ijt-1} + X_{ijt}\beta + NBPTS_j\delta + \epsilon_{ijt} \quad (3)$$

In Eq. (3), $NBPTS_j$ is a measure of teacher performance on the NBPTS assessment. We measure teacher outcomes in several different ways to produce different comparisons of teacher effectiveness.

In our most basic models, $NBPTS_j$ indicates that teacher j passes the National Board assessment on the first attempt. These regressions estimate the average difference in effectiveness between teachers who pass the assessment on the first attempt and other, initially unsuccessful NBPTS applicants. The estimates from these regressions may differ from those estimated with the entire sample of teachers above for two reasons. First, applicants for NBPTS certification, whether successful or unsuccessful, may be more or less effective than the average non-applicant. If NBPTS applicants are more effective than the average non-NBCT, then differences in value-added by certification status may be smaller within the sample of applicants than for the population of teachers as a whole. Second, initially unsuccessful applicants may reapply to the board for certification, so some of the NBCTs we observe in Section III initially failed their assessment.²⁵

²⁴ An additional concern is whether to include teachers who have not submitted assessment results. Some studies have included all teachers with indicators for having submitted an assessment. This may improve efficiency for the student- and classroom-level regressors, but point estimates are generally biased if assessment results are correlated with student and classroom characteristics (Jones, 1996). We therefore limit our sample to teachers with assessment outcomes.

²⁵ In the Washington data, we observe a 60% first-time pass rate and an 83% three-year pass rate. These numbers are higher than those reported nationally (Committee on Evaluation of Teacher Certification by the National Board for Professional Teaching Standards, 2008). However, among a sample of North Carolina teachers, which is another state with a large population of NBCTs, Goldhaber and Hansen (2008) find a first-time passing rate of 54% and an eventual passing rate of about 75%, which are roughly consistent with the patterns we observe. In the analytical samples, the pass rates are even higher: 65-75% for initial applicants and 85%-95% overall.

Therefore, we also include models with indicators for whether the teacher subsequently passes on a retake. These models compare initially successful applicants and those who pass on retakes to those who never obtain certification.

While the NBPTS certification decisions are binary, the underlying assessment process may contain additional information about teacher effectiveness. We therefore estimate models where $NBPTS_j$ is the teacher's assessment score. We standardize the NBPTS scores against the distribution of first-time assessments so that the estimated coefficients measure the difference in student achievement associated with a one standard deviation difference in NBPTS assessment scores. As with the binary passing indicator, teachers may retake portions of the NBPTS assessment and the first score does not correspond to the final certification decision for all teachers. We therefore estimate models that include both the initial score and the maximum score for each candidate. Suppose we have two candidates who both receive the same score and fail their first attempt but receive different scores on their second attempt. If teacher performance on the retake reflects differences in teacher effectiveness, we should observe a relationship between the final score and student achievement even after controlling for the first score. In other words, these regressions test whether the difference between the initial and final candidate scores adds any additional information about teacher effectiveness.

We present the results for differences in effectiveness by assessment outcomes in **Table 5**. In elementary classrooms, teachers who initially pass the NBPTS assessment are 0.06 standard deviations more effective than those who fail in teaching math and 0.05 standard deviations more effective in teaching reading. When we add indicators for subsequently passing the NBPTS assessment, we find that elementary teachers are approximately 0.09 standard deviations more effective than those who never pass. These latter effects are approximately the same size as those estimated by Goldhaber and Anthony (2007) and somewhat smaller than the experimental estimates reported by Cantrell et al. (2008). In terms of annual learning gains, our estimates suggest that the differences in effectiveness by initial performance on the NBPTS assessment correspond to about 4.5 weeks of learning.²⁶ When we additionally consider teachers who pass the NBPTS assessment after initially failing, we only find evidence that teachers who pass on a retake are more effective than those who never pass in reading.

In Panel B, we show results for middle school classrooms. Interestingly, we do not find that middle school teachers who initially pass National Board assessments are more effective than those who fail, although the effect is statistically significant at the 10% level for mathematics teachers. We find a difference of 0.06 standard deviations in math and 0.03 in reading classrooms, although neither of the coefficients is statistically significant. Adding indicators for passing on a subsequent administration does little to change these estimates. However, given the relatively smaller samples of middle school

²⁶ This conversion uses the findings from Bloom et al. (2008) and is discussed in footnote 21.

applicants and the high pass rates of the sample of teachers matched to classrooms, the estimated contrasts are generally imprecisely estimated.

Next, we consider teacher effectiveness by the initial score on the National Board assessment. We replace the indicator for passing the assessment in Eq. (3) with teachers' total assessment scores. Across subjects and school levels, we find that a one standard deviation difference on the National Board assessment score corresponds to an approximately 0.04-0.05 standard deviations difference in student achievement.²⁷ The results for mathematics are smaller than the experimental estimates from Cantrell et al. (2008) but similar to the non-experimental results estimated on a larger sample of teachers, while the reading results are similar to both sets of estimates. When we include teachers' maximum scores on the NBPTS assessment, we find little evidence that subsequent scores add additional explanatory power for predicting student achievement. In mathematics, the coefficient on the maximum score is small and statistically insignificant for both grade levels; in reading, the coefficients are larger but we do not find statistically significant evidence that they add additional information beyond what is contained in the first score.

To further explore the relationship between NBPTS assessment scores and student achievement, we additionally estimate models using quintiles of NBPTS assessment scores instead of a linear specification. We plot the coefficients for the lowest and highest two quintiles by subject and grade level in **Figure 1** (the middle quintile is the omitted group). A few interesting non-linearities are apparent from the figures. First, in no sample are the coefficients on the two lowest quintiles of performance jointly or individually statistically significantly different than the middle quintile of performance. In the elementary school sample, we find that the highest two quintiles of performance have similar average student achievement effects, which is consistent with the diminishing marginal effects found by Cantrell et al. (2008).²⁸ On the other hand, we find evidence in the middle school grades that teachers in the highest performance quintile are producing significantly higher student achievement effects. The highest quintile outperforms the fourth quintile by 0.10 student standard deviations in middle school math classrooms and 0.06 student standard deviations in middle school reading classrooms. Both of these differences are statistically significant at the 0.01 level.

To give some sense of the magnitude of these findings, it may be helpful to consider the additional variation in student achievement explained by the National Board assessment. We therefore estimate teacher and classroom random effects models that include controls for teacher experience on the sample of NBPTS applicants both with and without the final candidate assessment score. Without the final assessment score, we estimate the variance of teacher effectiveness among National Board applicants is 0.022

²⁷ We standardize all NBPTS assessment scores against the distribution of first-time assessment results across all certificates.

²⁸ The differences in average effectiveness are not statistically significant in either subject.

in elementary math, 0.015 in elementary school reading, 0.025 in middle school math, and 0.007 in middle school reading. Adding the final score to the value-added models explains about 4-5% of the variance in teacher effectiveness in mathematics, about 8% of the variance of teacher effectiveness in elementary reading, and about 11% of the variance in middle school reading. For comparison, Rockoff et al. (2011) consider several non-traditional measures of pre-service teacher quality and find that they explain about 10% of the variation in future teacher effectiveness.

We next break out the performance of National Board candidates by certificate type. For these regressions, we estimate teacher effectiveness using the sample of teachers who apply for the given certificate and are teaching in a related classroom. The results of these regressions are in **Tables 6** and **7**. The estimates are less precise than those that aggregate across certificate types, but produce generally consistent results. Teachers who pass the MC/Gen assessment on the first attempt are about 0.06 standard deviations more effective teaching math and 0.03 standard deviations more effective in teaching reading than those who initially fail, although the reading result is only statistically significant at the 0.10 level. In both cases, teachers who pass the assessment on either the first attempt or a retake are more effective than those who never pass. Results are somewhat larger for the EMC/LRLA assessment. Teachers who pass on the first attempt are about 0.16 standard deviations more effective in both math and reading. As with the results with the aggregated certificates, we find evidence that the first score predicts student achievement, but do not find consistent evidence that the maximum score adds any additional information.

We present the estimates for the Early Adolescence certificates using the middle school data in Table 7. In Table 5, nearly 75% of the applicants in the middle school math sample apply for certification in something other than EA/Math. When we limit the sample of teachers to those who apply for EA/Math certification, we find that first-time passers are about 0.08 standard deviations more effective than those who initially fail. When we split the sample of first-time unsuccessful applicants by their ultimate certification status, we do not find any statistically significant differences by certification outcome. This is due to the fact that the point estimates for the group of teachers who pass on a second attempt is actually negative; however, the group of teachers who never earn certification is only 31 and all of the point estimates are imprecisely estimated. For the EA/ELA certificate, we do not find statistically significant differences in teacher effectiveness by the assessment outcomes. As with the EA/Math certificate, however, the group of applicants who never pass is small and coefficients are imprecise. As with the other certificates, when we instead look at overall performance, we do find that continuous measures of performance predict student achievement for both certificates. For the EA/Math assessment, we estimate that a one standard deviation differences in assessment scores predicts about a 0.07 standard deviation increase in student

achievement. For the EA/ELA assessment, the estimated coefficient is similar to those observed in the larger sample of applicants.

Optimal Weights for Value-Added Prediction

The National Board assessment comprises ten separate exercises, and these are judged separately and then aggregated to obtain the final scale score. The assessment score used by NBPTS reflects their professional judgment of the relative importance of various teacher characteristics and skills and may not provide the optimal prediction of teacher quality among all possible combinations of the NBPTS assessment subscores (Cantrell et al., 2008). We therefore attempt to use the assessment information to better predict teacher value-added. An important caveat of the following analysis is that student mastery of the skills reflected by standardized assessments is only one responsibility of teaching. While recent research suggests that value-added reflects important contributions to both short- and long-term student outcomes, other research suggests that teachers may also make important contributions to higher-order analytical and non-cognitive skills that may not be well captured by value-added (Chetty et al., 2014a, 2014b; Jackson, 2012; Papay, 2010).

In order to estimate the optimal weights for value-added prediction, we replace the final NBPTS assessment score in Eq. (3) with the average assessment score for each of the four NBPTS exercise types (student work, instructional analysis, documented accomplishments, and assessment center exercises).^{29,30} We then form the optimal weights for value-added prediction by standardizing the regression coefficients to sum to one. We estimate the standard errors using the delta method.

We display the optimal weights by subject in **Table 8**. In column 2, we display the current weights for each exercise type. The final score weights the student work exercise at 0.16, the two instructional analysis exercises at 0.32 (0.16 each), the documented accomplishments exercise at 0.12, and the six assessment center exercises at 0.40 (0.40/6 each). Our estimated optimal prediction weights are in columns 3 and 4. In column 3, the optimal value-added prediction weights include the MC/Generalist and EA/Mathematics certificates. In column 4, the weights include the MC/Generalist, EA/English Language Arts, and EMC/Literacy, Reading, and Language Arts certificates. The optimal weights differ by subject, but both suggest greater weight should be placed on the documented accomplishments portfolio entries. However, the results differ across subjects for the other exercise types and all the weights are imprecisely estimated.

A more difficult question is whether adoption of the suggested weights is likely to improve the certification decision process. A simple way to check the performance of the

²⁹ We additionally constrain the coefficients on the score variables to be positive, although negativity of the coefficients on exercise type averages is not a concern in the present context.

³⁰ We found our sample sizes to be too small to produce reliable estimates of optimal weights when we treated each of the assessment exercises separately.

reweighted estimator is to compute the additional proportion of the variance in student achievement predicted by the reweighted assessment scores compared to the NBPTS-determined weights. These results suggest that the reweighted assessment scores provide better estimates of the observed teacher value-added for all the assessments. We find that the reweighted assessment scores improve the R^2 in our student achievement regression by about 0.0002 in math and 0.0001 in reading. We found above that the NBPTS assessment scores explained about 5-10% of the unobserved teacher component; these results suggest that the reweighted assessment scores explain nearly an additional 1% of unobserved teacher effectiveness.

However, this approach is likely to produce an optimistic view of the reweighted scores since we assess the fit of the model using the same teachers used to generate the weights. In other words, if we obtained a new sample of teachers with NBPTS assessment scores and applied the weights derived from our sample, we would expect to find a greater prediction error making out-of-sample predictions. We therefore use a cross-validation approach to assess whether our reweighting procedure creates better predictions of teacher value-added than the original weights. We implement the cross-validation procedure by randomly dividing each of the teacher samples into 10 nearly equally sized subsamples. For each subsample, we re-estimate the optimal weights using the students assigned to teachers in the remaining nine samples and then calculate the reweighted assessment score for the selected subsample. We then regress student achievement on the control variables and the reweighted assessment score using all subsamples. While we cannot assess whether our chosen weights would perform better on an entirely new sample of teachers, this resampling approach allows us to assess the *procedure* of choosing optimal weights. While we estimate better in-sample predictions for all assessments, the out-of-sample predictions do not perform as well. For both subjects, the reweighted assessment scores perform worse than the existing weights when making predictions of teacher effectiveness for an unseen population of teachers. While a larger sample of teachers may produce weights better aligned with teacher value-added, we conclude that the NBPTS weighting scheme provides reasonably good estimates of teacher value-added for the current choice of assessment exercises.

V. Policy Implications and Conclusions

In this study, we assess the relationship between teacher value-added and performance on the National Board for Professional Teaching Standards assessments. We find that teachers in Washington with the National Board certificate are between 0.01 – 0.05 standard deviations more effective than non-NBCTS, which is consistent with prior studies of NBCTs in North Carolina and Florida. For elementary teachers and middle school reading teachers, we find differences in effectiveness of about 0.01-0.02 standard deviations. In middle school math, NBCTs are about 0.05 standard deviations more effective than non-NBCTs. The differential result for middle school math classrooms

appears to be driven by the larger gap in average effectiveness between non-NBCTs and NBCTs certified under the EA/Math assessment.

Comparisons to educational benchmarks suggest that these differences may be of educational significance. Results from nationally normed tests suggest that the differences in teacher effectiveness for NBCTs may correspond to approximately 1-2 weeks of additional learning in elementary classrooms and middle school reading classrooms and nearly 1.5 months of additional learning in middle school math classrooms (Bloom et al., 2008). While estimates of the returns to teaching experience vary, the elementary and middle school reading results are approximately equal to 15-35% of the return to the first five years of teaching experience. The middle school mathematics results suggest that the effectiveness of NBCTs relative to non-NBCTs is about 50-75% of the return to the first five years of experience (Atteberry et al., 2013; Harris and Sass, 2011; Wiswall, 2013).

We further find that performance on the National Board assessments predicts student achievement, although this relationship varies across the different certificates offered by NBPTS. A one standard deviation difference in assessment scores appears to correspond to a difference of about 0.04-0.05 standard deviations in student achievement across all levels and subjects we consider, which corresponds to about 3-5 weeks of student learning gains. However, it appears that there may be important nonlinearities in the relationship between the assessment score and student achievement and we find some evidence that teachers in the top 40% of the NBPTS assessment score distribution produce substantially greater learning gains than those in the bottom 60%. Given the sample size of teachers available for this study, re-weighted composite scores designed to best predict value-added do not outperform the existing set of weights. Finally, we find some evidence that the retesting procedures of the NBPTS weaken the assessment's ability to differentiate between more and less effective teachers. For elementary and middle school reading teachers, we find no evidence that NBCTs who initially failed the NBPTS assessment but earned certification on a subsequent sitting are more effective than non-NBCTs. Notably, this result does not hold for middle school math teachers. Among applicants for NBPTS certification, comparisons of teachers who initially fail and subsequently pass to those who never pass are complicated by small sample sizes and produce more ambiguous results.

Over the past 10 years, Washington has revised its compensation policies surrounding National Board teachers and has dramatically increased the number of NBCTs in the state. Our analyses suggest that the teachers licensed in this time period are more effective than the average non-NBCT in the state. While our study does not speak to the policy effectiveness of any particular certification policy, we do find that NBCTs in high-poverty schools, who have received an additional bonus since 2008, are at least as effective relative to their colleagues than teachers in other schools.

A number of states are experimenting with policies aimed at improving the recruitment and retention of effective teachers. Often these involve financial incentives

for particular groups of teachers. Observable measures of teacher effectiveness are therefore an important prerequisite for such policies. The credential offered by the National Board for Professional Teaching Standards serves this role in 24 states as well as in other individual school districts (Exstrom, 2011). While our results provide only a descriptive analysis of the effectiveness of NBCTs, and do not indicate the effectiveness of any particular compensation policy, they do suggest that the teachers targeted by these incentives are likely on average more effective than the population of teachers as a whole. The overall efficacy of policies that incentivize NBCTs for improving student outcomes, however, is much harder to assess and there is little direct evidence on their impact. In particular, such policies rely on the sensitivity of teacher labor supply decisions to financial incentives and the effects of improved teacher recruitment and retention on student outcomes. A number of studies have found that teachers respond to financial incentives in deciding where to work or whether to leave the profession (Clotfelter et al., 2008; Dee and Wyckoff, 2013). Beyond any potential improvements in teacher staffing, reduced turnover may also directly affect student achievement (Ronfeldt et al., 2013). Although there is little empirical evidence of NBCT spillover effects, the effects of reduced turnover may be particularly salient for NBCTs given their high reported participation in leadership activities (Loeb et al., 2006). Nonetheless, there is little direct evidence on whether such incentive policies improve student outcomes. This includes both students' performance on standardized assessments as well as in other important domains. There is some evidence that teacher effects on non-tested outcomes may not be highly correlated with teacher value-added, but there is little evidence on the effects of credentials like the NBPTS on other student outcomes (Jackson, 2012). Further research is needed on the effects of these policies on teacher staffing and their implications for a variety of important student outcomes.

References

- Aaronson, D., Barrow, L., and Sander, W. (2007). Teachers and student achievement in the Chicago Public high schools. *Journal of Labor Economics*, 25(1): 95-135.
- Atteberry, A., Loeb, S., and Wyckoff, J. (2013). Do first impressions matter? Improvement in early career teacher effectiveness. *CALDER Working Paper*, 90: 1-51.
- Bloom, H. S., Hill, C. J., Black, A. R., and Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, 1(4): 289–328.
- Burke, M. A, & Sass, T. R. (2013). Classroom Peer Effects and Student Achievement. *Journal of Labor Economics*, 31(1), 51–82.
- Cantrell, S., Fullerton, J., Kane, T. J., and Staiger, D. O. (2008). National board certification and teacher effectiveness: Evidence from a random assignment experiment. *National Bureau of Economic Research Working Paper Series*, 14608.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, 104(9), 2593–2632.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, 104(9), 2633–2679.
- Chingos, M. M. and Peterson, P. E. (2011). 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): 449-465.
- Clotfelter, C. T., Glennie, E., Ladd, H., & Vigdor, J. (2008). Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina. *Journal of Public Economics*, 92(5-6), 1352–1370.
- Clotfelter, C. T., Ladd, H., & Vigdor, J. (2006). Teacher-student matching and the assessment of teacher effectiveness. *The Journal of Human Resources*, 41(4), 778–820.
- Clotfelter, C. T., Ladd, H., and Vigdor, J. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26(6): 673-682.
- Clotfelter, C. T., Ladd, H., and Vigdor, J. (2010). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources*, 45(3): 655-681.
- Committee on Evaluation of Teacher Certification by the National Board for Professional Teaching Standards. (2008). *Assessing accomplished teaching: Advanced-level certification programs*. (M. W. Hakel, J. A. Koenig, & S. W. Elliott, Eds.). Washington, D.C.: Board on Testing and Assessment, Center for Education, National Research Council.
- Dee, T., & Wyckoff, J. (2013). *Incentives, selection, and teacher performance: Evidence from IMPACT* (No. 19529). Cambridge, MA: National Bureau of Economic Research.

- Duflo, E., Dupas, P., & Kremer, M. (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*, 101(5), 1739–1774.
- Exstrom, M. (2011). National Board for Professional Teaching Standards certification: what legislators need to know. Technical report, National Conference of State Legislatures, Denver, Colo.
- Goldhaber, D. (2006). National Board Teachers Are More Effective, But Are They in the Classrooms Where They're Needed the Most? *Education Finance and Policy*, 1(3), 372–382.
- Goldhaber, D. and Anthony, E. (2007). Can teacher quality be effectively assessed? National Board certification as a signal of effective teaching. *Review of Economics and Statistics*, 89(1): 134-150.
- Goldhaber, D. and Hansen, M. (2013). Is it just a bad class? Assessing the long-term stability of estimated teacher performance. *Economica*, 80(319): 589–612.
- Goldhaber, D. D., Brewer, D. J., and Anderson, D. J. (1999). A three-way error components analysis of educational productivity. *Education Economics*, 7(3): 199-208.
- Goldhaber, D., Lavery, L., and Theobald, R. (2014). “Uneven Playing Field? Assessing the Teacher Quality Gap Between Advantaged and Disadvantaged Students.” *Educational Researcher*, in press.
- Harris, D. N., & Sass, T. R. (2009). The effects of NBPTS-certified teachers on student achievement. *Journal of Policy Analysis and Management*, 28(1), 55–80.
- Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95(7-8), 798–812.
- Humphrey, D. C., Koppich, J. E., & Hough, H. J. (2005). Sharing the wealth: National Board Certified teachers and the students who need them most. *Education Policy Analysis Archives*, 13(18).
- Ishii, J., & Rivkin, S. G. (2009). Impediments to the estimation of teacher value-added. *Education Finance and Policy*, 4(4), 520–536.
- Jackson, C. K. (2012). Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina. *National Bureau of Economic Research Working Paper Series*, 18624.
- Jackson, C. K. (2014). Teacher quality at the high school level: The importance of accounting for tracks. *Journal of Labor Economics*, 32(4), 645–684.
- Jacob, B. A. and Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1): 101-135.
- Johnson, M., Lipscomb, S., & Gill, B. (2014). Sensitivity of teacher value-added estimates to student and peer control variables. *Journal of Research on Educational Effectiveness*, forthcoming.
- Jones, M. P. (1996). Indicator and Stratification Methods for Missing Explanatory Variables in Multiple Linear Regression. *Journal of the American Statistical Association*, 91(433), 222-230.
- Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City. *Economics of Education Review*, 27(6), 615–631.

- Kane, T. J., Taylor, E. S., Tyler, J. H., and Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources*, 46(3): 587-613.
- Koppich, J. E., Humphrey, D. C., & Hough, H. J. (2007). Making use of what teachers know and can do: Policy, practice, and National Board Certification. *Education Policy Analysis Archives*, 15(7), 1–30.
- Lavy, V., Paserman, M. D., & Schlosser, A. (2012). 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), 208–237.
- Lefgren, L. (2004). Educational peer effects and the Chicago Public Schools. *Journal of Urban Economics*, 56(2), 169–191.
- Loeb, H., Efers, A. M., Plecki, M. L., Ford, B., & Knapp, M. S. (2006). *National Board Certified Teachers in Washington State: Impact on professional practice and leadership opportunities*. Seattle, WA: Center for Strengthening the Teaching Profession.
- National Board of Professional Teaching Standards (2010). Profiles in excellence: Washington state. Technical report.
- National Board of Professional Teaching Standards (2012). State profile: Washington. Technical report, Arlington, VA.
- National Board for Professional Teaching Standards. (2014a). *2014 state rankings by new number of National Board certified teachers*. Arlington, VA: National Board for Professional Teaching Standards.
- National Board for Professional Teaching Standards. (2014b). *2014 state rankings by total number of National Board certified teachers*. Arlington, VA: National Board for Professional Teaching Standards.
- Pearlman, M. (2008). The design architecture of NBPTS certification assessments. In Ingvarson, L. and Hattie, J., editors, *Assessing teachers for professional certification: The first decade of the National Board for Professional Teaching Standards*, number 11 in *Advances in Program Evaluation*, pages 55–91. JAI Press, Bingley, UK.
- Protik, A., Walsh, E., Resch, A., Isenberg, E., & Kopa, E. (2013). Does tracking of students bias value-added estimates for teachers? Presented at the Association of Education Finance and Policy Conference, New Orleans, LA.
- Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2): 417-458.
- Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy*, 6(1), 43–74.
- Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How Teacher Turnover Harms Student Achievement. *American Educational Research Journal*, 50(1), 4–36.
- Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy*, 4(4): 537-571.
- Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, 125(1): 175-214.
- Sato, M., Wei, R. C., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of National Board Certification. *American Educational Research Journal*, 45(3), 669–700.

- Snyder, T. D. and Dillow, S. A. (2013). Digest of education statistics 2011. Technical Report 2014-015, National Center for Education Statistics, Washington, D.C.
- Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics*, 100, 61–78.

Table 1. Summary Statistics

	Elementary		Middle School Math		Middle School Reading	
	All (1)	NBCT (2)	All (3)	NBCT (4)	All (5)	NBCT (6)
Math post-test	0.007 (0.998)	0.086 (1.024)	0.005 (0.995)	0.221 (1.022)		
Reading post-test	0.008 (0.997)	0.070 (1.004)			0.052 (0.966)	0.161 (0.948)
Math pre-test	0.006 (0.997)	0.054 (1.013)	0.009 (0.992)	0.183 (1.012)	0.037 (0.984)	0.133 (0.992)
Reading pre-test	0.003 (0.999)	0.037 (1.004)	-0.003 (0.998)	0.137 (0.984)	0.057 (0.960)	0.158 (0.947)
Female	0.492 (0.500)	0.492 (0.500)	0.494 (0.500)	0.495 (0.500)	0.501 (0.500)	0.503 (0.500)
American Indian	0.020 (0.139)	0.015 (0.122)	0.015 (0.123)	0.010 (0.101)	0.016 (0.124)	0.011 (0.106)
Asian/Pacific Islander	0.085 (0.279)	0.106 (0.307)	0.087 (0.281)	0.112 (0.316)	0.086 (0.280)	0.099 (0.299)
Black	0.048 (0.213)	0.044 (0.205)	0.043 (0.203)	0.038 (0.192)	0.041 (0.197)	0.037 (0.189)
Hispanic	0.172 (0.377)	0.177 (0.382)	0.172 (0.378)	0.166 (0.372)	0.173 (0.378)	0.174 (0.379)
White	0.631 (0.483)	0.601 (0.490)	0.632 (0.482)	0.624 (0.484)	0.634 (0.482)	0.627 (0.484)
Multiracial	0.043 (0.203)	0.056 (0.231)	0.050 (0.218)	0.049 (0.217)	0.051 (0.220)	0.051 (0.221)
Learning disabled	0.062 (0.240)	0.066 (0.248)	0.054 (0.226)	0.037 (0.189)	0.042 (0.201)	0.031 (0.173)
Gifted	0.050 (0.218)	0.070 (0.254)	0.073 (0.260)	0.092 (0.289)	0.075 (0.263)	0.105 (0.307)
Limited English proficient	0.066 (0.247)	0.075 (0.264)	0.038 (0.192)	0.037 (0.189)	0.031 (0.173)	0.031 (0.173)
Special education	0.125 (0.331)	0.130 (0.336)	0.096 (0.295)	0.071 (0.257)	0.078 (0.269)	0.061 (0.239)
Free/reduced-price lunch	0.447 (0.497)	0.454 (0.498)	0.432 (0.495)	0.402 (0.490)	0.427 (0.495)	0.408 (0.491)
Honors course			0.042 (0.202)	0.045 (0.207)	0.088 (0.283)	0.108 (0.311)
Remedial course			0.012 (0.109)	0.006 (0.077)	0.008 (0.088)	0.007 (0.086)
N	742158	49430	572102	61282	496458	63909

Table 2. Effectiveness of Board Certified Teachers (Elementary School Classrooms)

	Math			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Any Certificate						
NBCT	0.035*** (0.009)	0.018*** (0.007)	0.017* (0.009)	0.027*** (0.007)	0.016*** (0.006)	0.007 (0.008)
N	742158	742158	329360	742158	742158	329360
Panel B. Individual Certificates						
MC/GEN	0.034*** (0.010)	0.017** (0.008)	0.018* (0.010)	0.026*** (0.008)	0.011* (0.007)	0.002 (0.008)
EMC/LRLA	0.047** (0.019)	0.024* (0.014)	0.043** (0.020)	0.031** (0.014)	0.026** (0.011)	0.026 (0.016)
Other cert	0.019 (0.024)	0.015 (0.018)	-0.028 (0.024)	0.031 (0.019)	0.031** (0.015)	0.010 (0.022)
N	742158	742158	329360	742158	742158	329360
Panel C. Passing Attempt						
NBCT first attempt	0.051*** (0.010)	0.030*** (0.008)	0.028*** (0.010)	0.033*** (0.008)	0.022*** (0.006)	0.011 (0.009)
NBCT retake	-0.013 (0.017)	-0.017 (0.012)	-0.017 (0.015)	0.010 (0.014)	0.000 (0.010)	-0.005 (0.012)
N	742158	742158	329360	742158	742158	329360
Cohort FE	N	Y	Y	N	Y	Y
Apparently Random Sample	N	N	Y	N	N	Y
Number of Teachers:						
NBCT	903	903	580	903	903	580
MC/GEN	592	592	401	592	592	401
EMC/LRLA	183	183	105	183	183	105
Other certificate	128	128	74	128	128	74
NBCT first attempt	661	661	422	661	661	422
NBCT retake	242	242	158	242	242	158

Notes: Models in Panel A regress student achievement on indicator for teacher's National Board certification status, cubic polynomials in prior achievement in math and reading, student sex, race and ethnicity, FRL eligibility, learning disabled status, and participation in special education, English language learning, or gifted programs. Models in Panel B replace the NBCT indicator with indicators for subject-specific certificates. Panel C replaces NBCT indicator with indicators for a teacher who is an NBCT and passed the assessment on the first attempt or passed the assessment on a subsequent attempt. Cohorts indicate school-grade-year cells. Apparently random sample includes schools without clear evidence of sorting determined as described in text. Counts of teachers give the number of unique teachers with each certificate in the analysis sample. Standard errors in parentheses are clustered by the teacher level in all equations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3. Effectiveness of Board Certified Teachers (Middle School Classrooms)

	Math			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Any Certificate						
NBCT	0.049*** (0.012)	0.052*** (0.009)	0.049*** (0.009)	0.019** (0.008)	0.012** (0.005)	0.012** (0.005)
N	572102	572102	572102	496458	496458	496458
Panel B. Individual Certificates						
EA/Math	0.055*** (0.013)	0.065*** (0.010)	0.061*** (0.010)			
EA/ELA				0.021** (0.009)	0.013** (0.006)	0.014** (0.006)
Other cert	0.011 (0.027)	-0.007 (0.016)	-0.007 (0.016)	0.015 (0.013)	0.010 (0.009)	0.009 (0.009)
N	572102	572102	572102	496458	496458	496458
Panel C. Passing Attempt						
NBCT first attempt	0.063*** (0.013)	0.057*** (0.010)	0.054*** (0.010)	0.027*** (0.009)	0.017*** (0.006)	0.018*** (0.006)
NBCT retake	0.009 (0.024)	0.037*** (0.013)	0.034** (0.013)	-0.005 (0.014)	-0.005 (0.011)	-0.006 (0.011)
N	572102	572102	572102	496458	496458	496458
Number of Teachers:						
NBCT	371	371	371	511	511	511
EA/MATH	226	226	226	11	11	11
EA/ELA	17	17	17	284	284	284
Other cert	153	153	153	227	227	227
NBCT first attempt	257	257	257	365	365	365
NBCT retake	114	114	114	146	146	146

Notes: Models in Panel A regress student achievement on indicator for teacher's National Board certification status, cubic polynomials in prior achievement in math and reading, student sex, race and ethnicity, FRL eligibility, learning disabled status, and participation in special education, English language learning, or gifted programs. Models in Panel B replace the NBCT indicator with indicators for subject-specific certificates. Panel C replaces NBCT indicator with indicators for a teacher who is an NBCT and passed the assessment on the first attempt or passed the assessment on a subsequent attempt. Cohorts indicate school-grade-year cells; tracks additionally stratify cohorts by honors and remedial status. Standard errors in parentheses are clustered by the teacher level in all equations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4. National Board Effects by Student Subgroup

	Math			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Elementary School Classrooms						
NBCT	0.019** (0.008)		0.016** (0.008)	0.015** (0.007)		0.013* (0.007)
NBCT * Gifted	0.025 (0.019)	0.053** (0.022)	0.026 (0.019)	0.015 (0.016)	0.030 (0.020)	0.015 (0.016)
NBCT * ELL	-0.015 (0.011)	-0.011 (0.011)	-0.019* (0.011)	-0.029** (0.012)	-0.028** (0.013)	-0.031** (0.012)
NBCT * SPED	0.015* (0.009)	0.018** (0.009)	0.015* (0.009)	0.005 (0.010)	0.007 (0.010)	0.005 (0.010)
NBCT * FRL	-0.009 (0.007)	-0.016** (0.006)	-0.014** (0.007)	0.004 (0.007)	-0.001 (0.007)	0.003 (0.007)
NBCT * Challenging Sch.			0.030* (0.016)			0.011 (0.014)
N	742158	742158	742158	742158	742158	742158
Classroom FE	N	Y	N	N	Y	N
Panel B. Middle School Classrooms						
NBCT	0.059*** (0.010)		0.059*** (0.011)	0.015** (0.006)		0.014** (0.007)
NBCT * Gifted	0.008 (0.017)	0.020 (0.019)	0.008 (0.017)	-0.012 (0.015)	0.016 (0.018)	-0.012 (0.015)
NBCT * ELL	-0.005 (0.016)	-0.006 (0.017)	-0.005 (0.016)	-0.002 (0.017)	-0.012 (0.016)	-0.004 (0.017)
NBCT * SPED	-0.040*** (0.013)	-0.048*** (0.013)	-0.040*** (0.013)	-0.010 (0.014)	-0.011 (0.015)	-0.010 (0.014)
NBCT * FRL	-0.010 (0.007)	-0.016*** (0.006)	-0.011* (0.007)	-0.003 (0.006)	-0.008 (0.006)	-0.004 (0.006)
NBCT * Challenging Sch.			0.003 (0.017)			0.009 (0.012)
N	572102	572102	572102	496458	496458	496458
Classroom FE	N	Y	N	N	Y	N

Notes: Results from regression of student achievement on indicator for teacher's National Board certification status and interactions with shown characteristics, cubic polynomials in prior achievement in math and reading, student sex, race and ethnicity, FRL eligibility, learning disabled status, and participation in special education, English language learning, or gifted programs. FRL = subsidized lunch eligibility; SPED = special education services; ELL = English language learner. Standard errors are clustered by the teacher level in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5. NBPTS Assessment Results and Teacher Effectiveness (All Certificates)

	Math				Reading			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Elementary School Classrooms								
Pass (first attempt)	0.062*** (0.019)	0.085*** (0.026)			0.045*** (0.017)	0.093*** (0.025)		
Pass (retake)		0.038 (0.031)				0.078*** (0.029)		
Score (first attempt)			0.039*** (0.010)	0.034 (0.022)			0.043*** (0.009)	0.017 (0.021)
Score (retake)				0.005 (0.021)				0.027 (0.020)
Number of observations	32614	32614	32614	32614	32614	32614	32614	32614
Number of teachers:								
In assessment sample	731	731	731	731	731	731	731	731
Pass on first attempt	507	507	507	507	507	507	507	507
Have a subsequent attempt	179	179	179	179	179	179	179	179
Pass on retake	142	142	142	142	142	142	142	142
Panel B. Middle School Classrooms								
Pass (first attempt)	0.057* (0.030)	0.036 (0.043)			0.033 (0.022)	0.033 (0.036)		
Pass (retake)		-0.040 (0.047)				0.001 (0.042)		
Score (first attempt)			0.047*** (0.015)	0.065** (0.029)			0.038*** (0.012)	0.022 (0.025)
Score (retake)				-0.019 (0.028)				0.017 (0.023)
Number of observations	24933	24933	24933	24933	27052	27052	27052	27052
Number of teachers:								
In assessment sample	244	244	244	244	332	332	332	332
Pass on first attempt	161	161	161	161	247	247	247	247
Have a subsequent attempt	57	57	57	57	72	72	72	72
Pass on retake	45	45	45	45	62	62	62	62

Notes: Regressions of student achievement on indicator for teacher's National Board certification result, cubic polynomials in prior achievement in math and reading, student sex, race and ethnicity, FRL eligibility, learning disabled status, and participation in special education, English language learning, or gifted programs. All models estimated on sample of teachers with NBPTS submissions in two school years prior to and following assessment. Standard errors are clustered at the teacher level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6. NBPTS Assessment Results and Teacher Effectiveness (Early/Middle Childhood Certificates)

	Math				Reading			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. MC/Generalist (Elementary School Classrooms)								
Pass (first attempt)	0.064*** (0.021)	0.072** (0.030)			0.036* (0.020)	0.087*** (0.030)		
Pass (retake)		0.013 (0.034)				0.081** (0.034)		
Score (first attempt)			0.037*** (0.010)	0.050** (0.022)			0.040*** (0.009)	0.006 (0.024)
Score (maximum)				-0.014 (0.020)				0.034 (0.022)
Number of observations	22682	22682	22682	22682	22682	22682	22682	22682
Number of teachers:								
In assessment sample	490	490	490	490	490	490	490	490
Pass on first attempt	329	329	329	329	329	329	329	329
Have a subsequent attempt	133	133	133	133	133	133	133	133
Pass on retake	105	105	105	105	105	105	105	105
Panel B. EMC/Literacy, Reading, and Language Arts (Elementary School Classrooms)								
Pass (first attempt)	0.158*** (0.054)	0.232*** (0.072)			0.155*** (0.040)	0.253*** (0.053)		
Pass (retake)		0.124 (0.096)				0.163** (0.070)		
Score (first attempt)			0.037* (0.020)	0.025 (0.059)			0.056*** (0.021)	0.070 (0.056)
Score (maximum)				0.012 (0.058)				-0.013 (0.057)
Number of observations	7202	7202	7202	7202	7202	7202	7202	7202
Number of teachers:								
In assessment sample	162	162	162	162	162	162	162	162
Pass on first attempt	131	131	131	131	131	131	131	131
Have a subsequent attempt	21	21	21	21	21	21	21	21
Pass on retake	17	17	17	17	17	17	17	17

Notes: Regressions of student achievement on indicator for teacher's National Board certification result, cubic polynomials in prior achievement in math and reading, student sex, race and ethnicity, FRL eligibility, learning disabled status, and participation in special education, English language learning, or gifted programs. All models estimated on sample of teachers with NBPTS submissions in two school years prior to and following assessment. Standard errors are clustered at the teacher level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7. NBPTS Assessment Results and Teacher Effectiveness (Early Adolescence Certificates)

	EA/Math (Middle School Math Classrooms)				EA/ELA (Middle School Reading Classrooms)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pass (first attempt)	0.077** (0.032)	0.051 (0.048)			0.042 (0.027)	0.028 (0.043)		
Pass (retake)		-0.049 (0.052)				-0.018 (0.051)		
Score (first attempt)			0.070*** (0.020)	0.108*** (0.035)			0.041*** (0.014)	0.025 (0.031)
Score (maximum)				-0.039 (0.036)				0.016 (0.027)
Number of observations	21897	21897	21897	21897	20085	20085	20085	20085
Number of teachers:								
In assessment sample	181	181	181	181	212	212	212	212
Pass on first attempt	115	115	115	115	160	160	160	160
Have a subsequent attempt	46	46	46	46	47	47	47	47
Pass on retake	35	35	35	35	39	39	39	39

Notes: Regressions of student achievement on indicator for teacher's National Board certification result, cubic polynomials in prior achievement in math and reading, student sex, race and ethnicity, FRL eligibility, learning disabled status, and participation in special education, English language learning, or gifted programs. All models estimated on sample of teachers with NBPTS submissions in two school years prior to and following assessment. Standard errors are clustered at the teacher level.

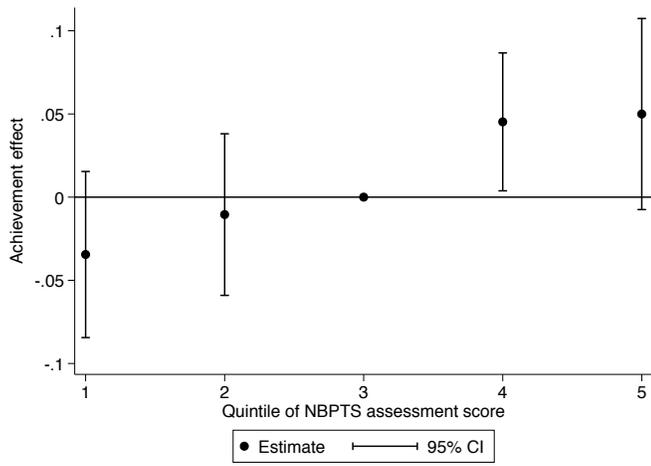
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8. Optimal Weights for Value-added Prediction

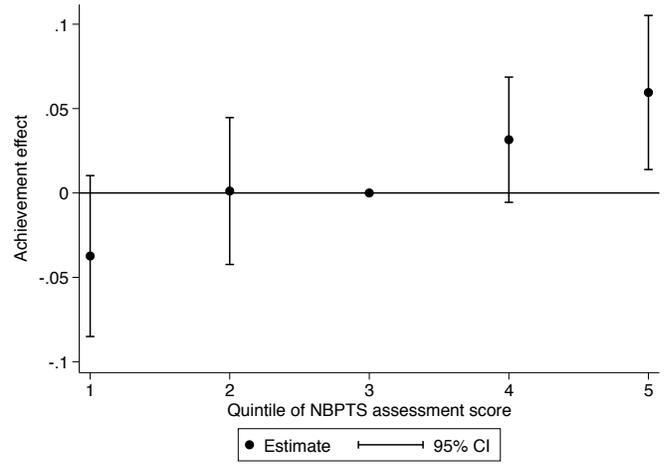
Type	No. Exercises	NBPTS Weight	Math	Reading
Student Work	1	0.16	0.08 (0.11)	0.20 (0.08)
Instructional Analysis	2	0.32	0.34 (0.14)	0.16 (0.10)
Documented Accomplishments	1	0.12	0.28 (0.12)	0.16 (0.07)
Assessment Center	6	0.40	0.30 (0.14)	0.48 (0.11)
Number of teachers			695	902
Number of student observations			46,064	51,970
In-sample change in R^2		0.0017/ 0.0012	0.0019	0.0013
Cross-validation change in R^2			0.0012	0.0011

Notes: Estimated optimal weights for value-added prediction by subject. Weights in column “NBPTS Weight” are current weights used in formulating the composite score. For math, included assessments are MC/Generalist and EA/ELA. For reading, included assessments are MC/Generalist, EA/ELA, and EMC/LRLA. In-sample change in R^2 gives improvement in R^2 of a regression of student achievement on controls from the addition of the weighted or reweighted NBPTS final score. Cross-validation change in R^2 is computed by the 10-fold cross-validation procedure described in the text. Totals may not sum to 1 due to rounding error. Standard errors of weights in parantheses are computed by the delta method and allow for clustering at the teacher level.

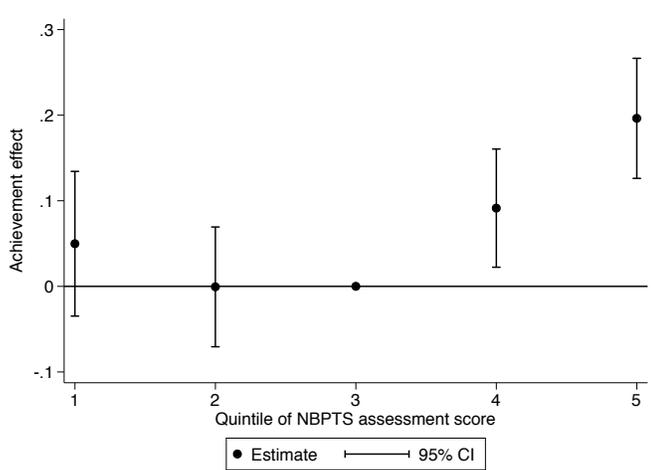
Figure 1. Student Achievement Effects by NBPTS Score Quintile



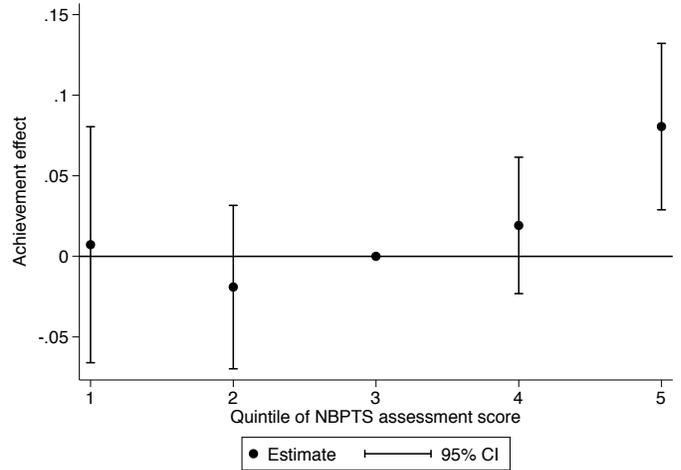
(a) Elementary Math



(b) Elementary Reading



(c) Middle School Math



(d) Middle School Reading