JOURNAL OF RESEARCH ON EDUCATIONAL EFFECTIVENESS
2016, VOL. 9, NO. 3, 233–258
http://dx.doi.org/10.1080/19345747.2015.1099768

INTERVENTION, EVALUATION, AND POLICY STUDIES

National Board Certification and Teacher Effectiveness: Evidence From Washington State

James Cowan[a] and Dan Goldhaber[a,b]

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.

KEYWORDS
teacher effectiveness
National Board for Professional Teaching Standards
teacher evaluation

Individual teachers have substantial influences on both immediate outcomes, such as standardized test scores and behavioral outcomes, and long-term outcomes, such as high school graduation, college attendance, and earnings (Aaronson, Barrow, & Sander, 2007; Chetty, Friedman, & Rockoff, 2014a, 2014b; Jackson, 2012; Nye, Konstantopoulos, & Hedges, 2004; Rivkin, Hanushek, & Kain, 2005). Yet, the credentials typically rewarded in the labor market, advanced degrees and experience, do not explain much of the variation in teacher quality (Goldhaber, Brewer, & Anderson, 1999; Goldhaber & Hansen, 2013; Harris & Sass, 2011; Kane, Rockoff, & Staiger, 2008). 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% 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

CONTACT James Cowan jecowan@gmail.com Center for Education Data and Research, 3876 Bridge Way North, Seattle, WA 98103, USA.
[a]Center for Education Data and Research, Seattle, Washington, USA
[b]American Institutes for Research, Washington, DC, USA

human capital purposes, the extant research on the effectiveness of National Board-Certified Teachers 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 is 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, Loeb, & Wyckoff, 2013; Bloom, Hill, Black, & Lipsey, 2008; Harris & 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.[1] We find that teachers who possess the National Board credential are about 0.02–0.05 standard deviations more effective than non-NBCTs with similar levels of experience in math. Our results are less robust for reading, but suggest that NBCTs are 0.01–0.02 standard deviations more effective than non-NBCTs in middle school classrooms and 0–0.02 standard deviations more effective in elementary classrooms. Comparing our results to the average achievement gains estimated from vertically aligned, nationally normed assessments, we estimate that NBCT effects correspond to about 4%–5% of normal annual learning gains at the elementary school level and for middle school reading and about 15% of annual learning gains in middle school math (Bloom et al., 2008). Finally, we find evidence that NBCT effectiveness differs based on whether the candidate gained certification on her first attempt or on a retake. 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. Except in middle school mathematics, we do not find evidence that teachers earning certification through a retake are more effective than non-NBCTs.

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. Because 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.[2] 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.

[1] 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.

[2] As of 2010, 39 states accept the NBPTS credential as a means of fulfilling state licensing or continuing education requirements (Exstrom, 2011).

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.[3] 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, Ladd, and Vigdor (2007) estimate statistically significant effects of 0.02–0.03 standard deviations for 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 nonrandom sorting of students to teachers in elementary school classrooms, Clotfelter, Ladd, and Vigdor (2006) find no evidence of NBCT effects in their most conservative models. Among high school teachers in North Carolina, Clotfelter, Ladd, and Vigdor (2010) find that NBCTs are about 0.05 standard deviations more effective than noncertified 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, Fullerton, Kane, and Staiger (2008) find no statistically significant differences between students in classrooms randomized to NBCTs and those in classrooms randomized to nonapplicants. 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. Although 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

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

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 etal., 2013; Bloom etal., 2008). Although 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.

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 & 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.[4] 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.[5] 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.[6] 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.[7] Overall, we matched 12,189 of the 12,309 NBPTS candidates (99%) to employment records in the S-275.

[4] Throughout this article, we refer to school years by the calendar year of the spring term.
[5] 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.
[6] 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.
[7] 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.

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 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.[8]

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.[9] 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.[10]

We present summary statistics for our analytical data set 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.

At the elementary level, the MC/Generalist certificate is by far the most common. In our sample, 7% of all classrooms and 71% 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% of all classrooms taught by an NBCT. For middle school students, the EA/Math and

8 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.

9 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.

10 Because 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 sixth grade.

Table 1. Summary statistics.

	Elementary		Middle school math		Middle school reading	
	All (1)	NBCT (2)	All (3)	NBCT (4)	All (5)	NBCT (6)
Math posttest	0.007	0.086	0.007	0.223		
	(0.998)	(1.024)	(0.994)	(1.021)		
Reading posttest	0.008	0.069			0.052	0.162
	(0.997)	(1.004)			(0.966)	(0.948)
Math pretest	0.006	0.053	0.011	0.185	0.037	0.135
	(0.997)	(1.013)	(0.991)	(1.012)	(0.984)	(0.993)
Reading pretest	0.003	0.036	−0.001	0.139	0.058	0.160
	(0.999)	(1.004)	(0.996)	(0.983)	(0.960)	(0.946)
Female	0.492	0.492	0.494	0.495	0.501	0.503
	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
American Indian	0.020	0.015	0.015	0.010	0.016	0.011
	(0.139)	(0.122)	(0.123)	(0.101)	(0.124)	(0.107)
Asian/Pacific Islander	0.085	0.106	0.087	0.113	0.086	0.100
	(0.279)	(0.307)	(0.282)	(0.316)	(0.280)	(0.299)
Black	0.048	0.044	0.043	0.038	0.041	0.037
	(0.213)	(0.206)	(0.203)	(0.192)	(0.198)	(0.190)
Hispanic	0.172	0.177	0.172	0.166	0.173	0.173
	(0.377)	(0.381)	(0.378)	(0.372)	(0.378)	(0.378)
White	0.631	0.601	0.632	0.624	0.633	0.627
	(0.483)	(0.490)	(0.482)	(0.484)	(0.482)	(0.484)
Multiracial	0.043	0.056	0.050	0.049	0.051	0.051
	(0.203)	(0.231)	(0.218)	(0.217)	(0.221)	(0.221)
Learning disabled	0.062	0.066	0.053	0.037	0.042	0.031
	(0.240)	(0.248)	(0.224)	(0.188)	(0.201)	(0.173)
Gifted	0.050	0.070	0.073	0.092	0.075	0.105
	(0.218)	(0.254)	(0.260)	(0.289)	(0.263)	(0.307)
Limited English proficient	0.066	0.076	0.038	0.037	0.031	0.031
	(0.247)	(0.264)	(0.192)	(0.188)	(0.173)	(0.173)
Special education	0.125	0.130	0.095	0.070	0.078	0.061
	(0.331)	(0.336)	(0.293)	(0.256)	(0.269)	(0.239)
Free/reduced-price lunch	0.447	0.455	0.432	0.402	0.427	0.407
	(0.497)	(0.498)	(0.495)	(0.490)	(0.495)	(0.491)
Honors course			0.043	0.045	0.088	0.109
			(0.202)	(0.207)	(0.284)	(0.311)
Remedial course			0.012	0.006	0.008	0.007
			(0.109)	(0.077)	(0.088)	(0.086)
N	742,124	49,450	570,533	61,184	492,800	63,679

EA/ELA certificates are the most common. Among all math classrooms, 9% are taught by an NBCT, and 7% are taught by a teacher with the EA/Math credential. In reading, NBCTs teach 11% of middle school classrooms, and teachers with an EA/ELA certificate teach nearly 7% of classrooms.

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:

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

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, participation in special education, English language learning, or gifted programs; we include in X_{jt} the teacher-year means of all of these variables.[11] 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. Because 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 Equation (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 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 & Chaplin, 2015; Kane, McCaffrey, Miller, & Staiger, 2013; Kane & 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, Walsh, Resch, Isenberg, & Kopa, 2013).

Even if value-added measures produce nearly unbiased predictions of future student achievement on average, it remains possible that biases in teacher effects are more substantial 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, such as 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, Paserman, & Schlosser, 2012; Lefgren, 2004). Although 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

[11] Using district-level data that permits better identification of discrete classrooms, Johnson, Lipscomb, and Gill (2015) 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.

distribution. For instance, higher-achieving students may benefit disproportionately from enrolling in classes with other high-achieving students (Burke & Sass, 2013; Duflo, Dupas, & Kremer, 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 reestimate our models with cohort effects on samples of schools for which there is little evidence of classroom sorting by observable student characteristics, that is, 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.[12] In 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.[13] Given that 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 et al. (2013) and include cohort-by-track fixed effects for our middle school sample.[14] 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 0.04 standard deviations more effective in math and 0.03 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.[15] The sample with apparently random assignment of students to classrooms includes about two thirds the NBCTs as the main analytical sample. When we limit the sample to balanced classrooms and include cohort fixed effects, the main NBCT effects are no longer statistically significant. In math, the point estimate is nearly identical and is significant at the 0.10 level.

[12] Our chi-square tests include indicators for whether the student scored above the median on each of the 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.

[13] 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 nonrandom assignment, we find that none of the models rejects the null hypothesis of no classroom effects at any conventional level.

[14] Jackson (2014) uses a finer designation of tracks at the high school level by using groups of students who take the same courses. Because our data set 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.

[15] 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.

Table 2. Effectiveness of board-certified teachers (elementary school classrooms).

	Math			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Any Certificate						
NBCT	0.037***	0.019***	0.017*	0.028***	0.017***	0.007
	(0.009)	(0.007)	(0.009)	(0.007)	(0.006)	(0.008)
N	742,124	742,124	329,345	742,124	742,124	329,345
Panel B. Individual Certificates						
MC/GEN	0.036***	0.018**	0.018*	0.026***	0.012*	0.002
	(0.010)	(0.008)	(0.010)	(0.008)	(0.007)	(0.008)
EMC/LRLA	0.050***	0.026*	0.043**	0.032**	0.027**	0.025
	(0.019)	(0.014)	(0.020)	(0.014)	(0.011)	(0.016)
Other cert	0.020	0.016	−0.028	0.032	0.032**	0.010
	(0.024)	(0.018)	(0.024)	(0.019)	(0.015)	(0.022)
N	742,124	742,124	329,345	742,124	742,124	329,345
Panel C. Passing Attempt						
NBCT first attempt	0.053***	0.032***	0.028***	0.034***	0.022***	0.011
	(0.010)	(0.008)	(0.010)	(0.008)	(0.006)	(0.009)
NBCT retake	−0.010	−0.016	−0.017	0.011	0.001	−0.005
	(0.017)	(0.012)	(0.015)	(0.014)	(0.010)	(0.012)
N	742,124	742,124	329,345	742,124	742,124	329,345
Cohort FE	N	Y	Y	N	Y	Y
Apparently random sample	N	N	Y	N	N	Y
Number of teachers:						
NBCT	904	904	580	904	904	580
MC/GEN	593	593	401	593	593	401
EMC/LRLA	183	183	105	183	183	105
Other cert	128	128	74	128	128	74
NBCT first attempt	661	661	422	661	661	422
NBCT retake	243	243	158	243	243	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.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

However, the result is less robust in reading, as the point estimate falls to 0.01 and is not significant at any conventional level.

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 find that these teachers are about 0.03 standard deviations more effective in reading than non-NBCTs; the point estimate is quite similar in math, but only statistically significant at the 0.10 level.

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

Table 3. Effectiveness of board-certified teachers (middle school classrooms).

	Math			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Any Certificate						
NBCT	0.051***	0.053***	0.050***	0.021***	0.014***	0.015***
	(0.012)	(0.009)	(0.009)	(0.007)	(0.005)	(0.005)
N	570,533	570,533	570,533	492,800	492,800	492,800
Panel B. Individual Certificates						
EA/Math	0.059***	0.067***	0.063***			
	(0.013)	(0.010)	(0.010)			
EA/ELA				0.023***	0.014**	0.015**
				(0.009)	(0.006)	(0.006)
Other cert	0.016	−0.001	−0.002	0.018	0.015*	0.014
	(0.027)	(0.016)	(0.016)	(0.013)	(0.009)	(0.009)
N	570,533	570,533	570,533	492,800	492,800	492,800
Panel C. Passing Attempt						
NBCT first attempt	0.064***	0.058***	0.055***	0.028***	0.019***	0.020***
	(0.013)	(0.010)	(0.010)	(0.008)	(0.006)	(0.006)
NBCT retake	0.013	0.039***	0.036***	−0.001	−0.001	−0.002
	(0.024)	(0.013)	(0.013)	(0.013)	(0.011)	(0.011)
N	570,533	570,533	570,533	492,800	492,800	492,800
Cohort FE	N	Y	N	N	Y	N
Track FE	N	N	Y	N	N	Y
Number of teachers						
NBCT	371	371	371	510	510	510
EA/MATH	218	218	218	11	11	11
EA/ELA	17	17	17	284	284	284
Other cert	153	153	153	226	226	226
NBCT first attempt	257	257	257	364	364	364
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.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

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 noncertified 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, 0.06–0.07 standard deviations more effective than noncertified teachers.[16] Overall,

[16] An alternative explanation for the difference between NBCT effects at the middle and elementary school level is that later cohorts of NBCTs are more effective than earlier cohorts and the middle school effect is a composition effect caused by the different coverage of elementary and middle school classrooms. To rule out this explanation, we reestimate the elementary regressions using only data from 2010 and later and find very similar results. Results are available from the authors upon request.

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 0.01 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 generally less effective than NBCTs who earn certification on their first attempt. When we stack the elementary and middle school data, we reject the hypothesis that the two groups are equally effective in both math and reading, although the result varies for individual subjects and grade levels.[19] However, middle school math teachers appear to be an exception: 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 at any conventional level.[20] Although 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.

The differential result for repeat applicants in middle school math appears to be driven by differences in the candidate samples across the grade levels and subjects. When we estimate models with indicators for whether a teacher has been a candidate for Board certification (Table A1), we generally find that unsuccessful applicants are less effective than the mean teacher. Although the results are not consistently statistically significant across specifications,

[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 within-teacher differences in effectiveness that are not statistically significant. These results are consistent with most of the prior results using student test score data and specifications with teacher fixed effects (Chingos & Peterson, 2011; Goldhaber & Anthony, 2007; Harris & 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.

[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 $F = 11.4$ ($p < 0.01$) for elementary math, $F = 3.7$ ($p = 0.05$) for elementary reading, and $F = 2.8$ ($p = 0.09$) 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 ($F = 9.7$; $p < 0.01$) and reading ($F = 5.9$; $p = 0.02$).

[20] The F statistic from the test of equality of the coefficients is $F = 1.5$ ($p = 0.22$) for middle school math.

we estimate that unsuccessful applicants are 0.03–0.07 standard deviations less effective than average at the elementary math level; 0–0.05 standard deviations less effective for elementary reading; and 0–0.06 standard deviations less effective for middle school reading. In each subject and grade level, at least two of the three specifications produce a statistically significant result at the 5% level. For middle school math, however, the point estimates are consistently about −0.01 standard deviations and all are statistically insignificant.

Overall, we find that certified teachers are more effective than noncertified 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 & Anthony, 2007) and Florida (Chingos & 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 Subgroups

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 teachers' student assessment skills (Goldhaber & Anthony, 2007; Sato, Wei, & 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, Koppich, & Hough, 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 (Chetty et al., 2014a; Clotfelter et al., 2007; Goldhaber, Lavery, & Theobald, 2015; Sass, Hannaway, Xu, Figlio, & Feng, 2012). Yet, Koppich, Humphrey, and Hough (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.

[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, respectively.

In order to better understand the effectiveness of NBCTs for disadvantaged students, we estimate NBCT effects separately for students participating in special programs. We estimate models such as Equation (1) that are fully interacted with a program participation indicator as follows:

$$A_{igjt} = A_{igjt-1}\rho_g + X_{igjt}\beta_g + NBCT_{jt}\delta_g + T_{jt}\gamma_g + \varepsilon_{igjt}. \quad (2)$$

In Equation (2), $g \, \varepsilon \, \{0, 1\}$ indicates whether the student belongs to the particular subgroup with nonparticipants defined as the baseline group. The coefficient δ_0 therefore gives the effect of NBCTs with nonparticipants, while δ_1 gives the effects of NBCTs for the particular student subgroup. 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 Equation (1), the regression models additionally include school-by-year-by-grade effects.

The difference in the estimated NBCT coefficients in Equation (2), $\delta_1 - \delta_0$, gives the average difference in achievement for students of the given subgroup relative to other students with an NBCT. A positive difference suggests that students of this particular subgroup have higher achievement than other students assigned to an 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 and that NBCTs therefore specialize in teaching these students. 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. For instance, 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., 2015).

In order to differentiate between these two possibilities, we additionally estimate Equation (2) with classroom fixed effects to control for any fixed teacher quality component.[22] The two NBPTS interaction terms (and all other interacted variables fixed at the classroom level) are collinear with the addition of classroom effects, so we drop the baseline NBCT indicator and only estimate the coefficient on the NBCT-subgroup interaction. The coefficients in these models therefore yield the interaction $\delta_1 - \delta_0$. Returning to the example above, if the difference in achievement (conditional on prior test scores and other covariates) between gifted and nongifted students is larger in NBCT classrooms than non-NBCT classrooms, we would conclude that NBCTs are relatively more effective at teaching gifted students. Because the classroom fixed effects remove generalized teaching effectiveness, we interpret the coefficients from these models as a test of whether NBCTs specialize in teaching certain groups of students.

The state incentive for NBCTs to teach in high-poverty schools may influence the effectiveness of NBCTs that are assigned to high-poverty students. We therefore conduct two

[22] Specifically, we control for teacher-by-track fixed effects, which may not uniquely identify classrooms in middle schools (Johnson et al., 2015).

additional analyses to test whether the incentive policy contributes to any observed subgroup differences. First, we estimate subgroup analyses treating the school eligibility for the CSB incentive as a "group." These models, which include school–grade–year fixed effects, therefore estimate the average difference in effectiveness between NBCTs and non-NBCTs in eligible and ineligible schools. As before, the coefficient δ_1 gives the difference in effectiveness between NBCTs and non-NBCTs for schools eligible for the CSB, while the interaction δ_1–δ_0 provides a test for whether the NBCT effect is the same in CSB and non-CSB schools. Second, to explore whether the subgroup effects can be explained by the different incentives for teachers to earn National Board certification in high-poverty schools, we reestimate Equation (2) using only schools ineligible for the high-poverty NBCT incentive. The coefficient δ_1 therefore provides an estimate of the effect of NBCTs for the given subgroup in non-CSB schools.

We present the results of the student-level heterogeneity regressions for elementary classrooms in Table 4 and for middle school classrooms in Table 5. Each of the subgroup analyses represents the results from a separate regression and we display estimates of the total (δ_1) and interaction (δ_1–δ_0) effects in each row. We focus first on the total effect of being assigned an NBCT rather than a non-NBCT in columns (1) and (5) and then the evidence on subgroup specialization for models with classroom fixed effects in columns (2) and (4). We include the interaction effects in columns (1) and (5) for comparison; however, recall that these estimates include both any specialization for particular subgroups and any variation in the effectiveness of NBCTs assigned to teach those subgroups.

Table 4. National Board effects by student subgroup (elementary school classrooms).

	Math				Reading			
	Cohort FE (1)	Class FE (2)	Non-CSB (3)	CSB (4)	Cohort FE (5)	Class FE (6)	Non-CSB (7)	CSB (8)
Gifted: Interaction	0.054**	0.048**	0.048*	0.055	0.004	0.034*	0.006	−0.066
	(0.025)	(0.022)	(0.026)	(0.096)	(0.023)	(0.020)	(0.024)	(0.106)
Gifted: Total effect	0.068***		0.058**	0.084	0.020		0.020	−0.044
	(0.024)		(0.025)	(0.096)	(0.023)		(0.024)	(0.107)
ELL: Interaction	−0.026*	−0.010	−0.022	−0.049**	−0.013	−0.022*	0.003	−0.038*
	(0.015)	(0.011)	(0.018)	(0.021)	(0.014)	(0.013)	(0.020)	(0.020)
ELL: Total effect	−0.007		−0.008	−0.007	0.003		0.017	−0.012
	(0.015)		(0.017)	(0.022)	(0.014)		(0.019)	(0.019)
SPED: Interaction	0.011	0.013	0.013	0.011	0.005	0.000	0.007	−0.001
	(0.010)	(0.009)	(0.011)	(0.020)	(0.011)	(0.010)	(0.013)	(0.021)
SPED: Total effect	0.028***		0.026**	0.045**	0.022**		0.021*	0.024
	(0.010)		(0.011)	(0.020)	(0.011)		(0.012)	(0.021)
FRL: Interaction	−0.012	−0.015**	−0.018**	−0.051**	0.009	−0.003	0.010	−0.013
	(0.009)	(0.007)	(0.009)	(0.022)	(0.009)	(0.007)	(0.009)	(0.025)
FRL: Total effect	0.011		0.002	0.029*	0.022***		0.021***	0.020
	(0.008)		(0.009)	(0.016)	(0.007)		(0.008)	(0.013)
CSB: Interaction	0.021				0.009			
	(0.017)				(0.014)			
CSB: Total effect	0.035**				0.023*			
	(0.015)				(0.012)			
N	742,124	724,142	637,033	105,091	742,124	742,124	637,033	105,091

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; CSB = Challenging Schools Bonus eligible. Standard errors are clustered by the teacher level in parentheses.
**p* < 0.10, ***p* < 0.05, ****p* < 0.01.

Table 5. National Board effects by student subgroup (middle school classrooms).

	Math				Reading			
	Cohort FE (1)	Class FE (2)	Non-CSB (3)	CSB (4)	Cohort FE (5)	Class FE (6)	Non-CSB (7)	CSB (8)
Gifted: Interaction	−0.021	0.021	−0.030	−0.020	0.010	0.019	0.014	0.035
	(0.021)	(0.020)	(0.025)	(0.034)	(0.018)	(0.018)	(0.021)	(0.029)
Gifted: Total effect	0.033*		0.026	0.020	0.025		0.028	0.058**
	(0.019)		(0.023)	(0.033)	(0.017)		(0.020)	(0.028)
ELL: Interaction	0.011	−0.015	−0.025	0.033	0.019	−0.012	0.009	0.008
	(0.020)	(0.015)	(0.029)	(0.025)	(0.023)	(0.017)	(0.028)	(0.030)
ELL: Total effect	0.063***		0.028	0.070***	0.031		0.020	0.028
	(0.020)		(0.030)	(0.026)	(0.023)		(0.028)	(0.031)
SPED: Interaction	−0.023*	−0.038***	−0.014	−0.038	−0.019	−0.019	−0.037**	0.025
	(0.013)	(0.013)	(0.016)	(0.027)	(0.014)	(0.015)	(0.016)	(0.028)
SPED: Total effect	0.029**		0.037**	0.001	−0.006		−0.022	0.038
	(0.013)		(0.015)	(0.025)	(0.014)		(0.016)	(0.028)
FRL: Interaction	0.007	−0.013**	0.005	0.027*	0.000	−0.007	−0.004	−0.001
	(0.009)	(0.006)	(0.010)	(0.015)	(0.008)	(0.006)	(0.009)	(0.016)
FRL: Total effect	0.056***		0.055***	0.047***	0.013*		0.010	0.021*
	(0.010)		(0.012)	(0.014)	(0.007)		(0.009)	(0.012)
CSB: Interaction	−0.013				0.006			
	(0.017)				(0.012)			
CSB: Total effect	0.040***				0.018*			
	(0.014)				(0.010)			
N	570,533	570,533	449,100	121,433	492,800	492,800	385,154	107,646

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; CSB = Challenging Schools Bonus eligible. Standard errors are clustered by the teacher level in parentheses.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

At the elementary level, we only find consistent evidence that NBCTs are more effective than non-NBCTs with special education students. The total effect of being assigned an NBCT for a special education student is 0.03 standard deviations in math and 0.02 standard deviations in reading. For both subjects, the total effect is similar to the average effect estimated in Table 2. We find that NBCTs are more effective in teaching math to gifted students than non-NBCTs by 0.07 standard deviations. We estimate that NBCTs are more effective in reading for FRL-eligible students by 0.02 standard deviations. The results suggest that NBCTs specialize in teaching gifted mathematics. We find that gifted students in NBCT classrooms outperform their peers by 0.05 standard deviations more than they do in non-NBCT classrooms. We also find evidence that NBCTs are relatively less effective with low-income children. We estimate that they underperform their classmates relative to low-income children in other classrooms by 0.02 standard deviations. We find little evidence of NBCT specialization in reading.

The results for middle school classrooms are displayed in Table 5. Given the greater efficacy of NBCTs for middle school math, we estimate positive total effects on assignment to an NBCT for ELL, special education, and FRL-eligible students, while the result for gifted students is statistically significant only at the 10% level. However, for reading, we find null effects of being assigned an NBCT for each of the groups considered. Despite the positive overall effects, we find that NBCTs appear to perform relatively worse with special education and low-income students. As at the elementary level, we find little evidence that NBCTs specialize with any particular group of students in reading.

Because the state incentive policy likely affects the distribution of NBCTs across student demographic groups and this may influence our findings, we consider teacher effectiveness separately by the school eligibility for the Challenging Schools Bonus (CSB). We begin by estimating the effect of an NBCT relative to other teachers in the same school for CSB-eligible and CSB-ineligible schools in the last row of columns (1) and (5). The interaction with the Challenging Schools Bonus indicator is statistically insignificant in both subjects and grade levels, which suggests that the difference in teacher effectiveness between NBCTs and non-NBCTs in challenging schools is similar to other schools. We also estimate models in columns (3)–(4) and (5)–(6) that estimate the subgroup effects separately for schools eligible for the Challenging Schools Bonus (CSB). The pattern of results is generally similar in both CSB and non-CSB schools. In particular, the point estimates for the effect of an NBCT for low-income students are similar to the average effects estimated in Tables 2 and 3, although only statistically significant for middle school math students.

Our estimates of the subgroup heterogeneity of NBCT effects are somewhat at odds with prior research by Goldhaber and Anthony (2007) that finds NBCTs appear to be more effective with FRL-eligible students. By contrast, our estimates suggest that NBCTs are relatively less effective in teaching FRL students in math and no more effective in reading. For elementary mathematics, we do not find any overall benefit for assigning FRL-eligible students to NBCTs. Similarly, for middle school math, we find that NBCTs perform relative less well with special education and low-income students, although the overall effect of being assigned an NBCT remains positive. The Washington incentive for NBCTs to teach in high-poverty schools does not appear to explain this result given that the patterns also hold in schools ineligible for the bonus. However, Goldhaber and Anthony (2007) do study teachers certified under the first-generation NBPTS assessment, which placed less emphasis on the assessment center exercises. It may be the case that the prior iteration of the NBPTS assessment placed more weight on teaching skills useful for low-income students. More generally, our results suggest that considering subgroup effects may be important when designing policies to help disadvantaged students. Our estimates suggest that, at least in math, the teachers targeted by the high-poverty incentive policy are not specialized in teaching high-poverty students. In fact, we find inconsistent evidence for the proposition that being assigned an NBCT has a positive effect on learning for low-income students.

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. Although 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 & Kraft, 2015; Rockoff, 2004). However, recent research also suggests that long-run teacher effects are not perfectly persistent across time (Chetty etal., 2014a; Goldhaber & 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 $t+2$.[23]

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:

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

In Equation (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 nonapplicant. 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 "Board Certification and Teacher Effectiveness" initially failed their assessment.[24] Therefore, we also include modes with indicators for whether the teacher subsequently

[23] 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.

[24] 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 (Hakel, Koenig, & Elliott, 2008). However, among a sample of North Carolina teachers, which is another state with a large population of NBCTs, Goldhaber and Hansen (2009) 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.

passes on a retake. These models compare initially successful applicants and those who pass on retakes to those who never obtain certification.

Although 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. Consequently, we 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. Finally, we test for nonlinearities in the relationship between NBPTS assessment scores and teacher effectiveness in two ways. We first regress student achievement on both an indicator for the binary pass/fail result and the continuous assessment score to test whether there is additional information about teacher effectiveness in the NBPTS results beyond the binary outcome. Second, we replace the continuous score measure with indicators for each quintile of the NBPTS assessment distribution.

We present the results for differences in effectiveness by assessment outcomes in Table 6. 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 0.09–0.10 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.[25] 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 applicants and the high pass rates of the sample of teachers matched to classrooms, the estimated contrasts are generally imprecisely estimated.

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

Table 6. National Board effects by student subgroup (middle school classrooms).

	Math					Reading				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. Elementary School Classrooms										
Pass (first attempt)	0.062***	0.086***			0.003	0.045***	0.097***			−0.042*
	(0.019)	(0.027)			(0.028)	(0.017)	(0.025)			(0.025)
Pass (retake)		0.040					0.083***			
		(0.032)					(0.029)			
Score (first attempt)			0.045***	0.029	0.044***			0.047***	0.011	0.065**
			(0.010)	(0.020)	(0.015)			(0.009)	(0.020)	(0.013)
Score (retake)				0.017					0.039**	
				(0.020)					(0.019)	
N	32,507	32,507	32,507	32,507	32,507	32,507	32,507	32,507	32,507	32,507
Panel B. Middle School Classrooms										
Pass (first attempt)	0.058*	0.037			−0.043	0.031	0.031			−0.020
	(0.030)	(0.043)			(0.044)	(0.022)	(0.037)			(0.028)
Pass (retake)		−0.039					−0.000			
		(0.047)					(0.042)			
Score (first attempt)			0.048***	0.061**	0.065***			0.037***	0.018	0.043***
			(0.015)	(0.028)	(0.022)			(0.012)	(0.024)	(0.016)
Score (retake)				−0.012					0.020	
				(0.027)					(0.023)	
N	24,847	24,847	24,847	24,847	24,847	26,928	26,928	26,928	26,928	26,928

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.

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

Next, we consider teacher effectiveness by the initial score on the National Board assessment. We replace the indicator for passing the assessment in Equation (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.[26] The results for mathematics are smaller than the experimental estimates from Cantrell et al. (2008) but similar to the nonexperimental 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 that subsequent scores add additional explanatory power for predicting student achievement only for elementary reading teachers. In mathematics, the coefficient on the maximum score is small and statistically insignificant for both grade levels.

We begin our assessment of nonlinearities in the relationship between the NBPTS assessment and student achievement by testing whether the binary passing indicator explains additional variation in achievement conditional on a linear function of the candidate's initial score. The results of these specifications are included in column (5) for math and in column (10) for reading. When we conduct this "horserace" of the passing threshold and linear score, we find that all of the information about teacher performance is contained in the composite score. The slope terms are generally similar to those without indicators for whether the teacher passed, albeit somewhat larger in middle school reading and middle school math. Each slope coefficient is statistically significant at the 5% level. On the other hand, the passing indicators are small or negative and statistically insignificant. Only the coefficient for elementary reading, which is negative, approaches statistical significance at the 5% level. Although the NBPTS standards for certification represent the consensus judgment of expert panels, these results appear consistent with prior research on the predictive validity of passing thresholds and continuous assessment outcomes for standardized licensure tests (Goldhaber, 2007) and teacher prescreening instruments (Rockoff, Loeb, & Wyckoff, 2011).

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 nonlinearities 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).[27] 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

[26]We standardize all NBPTS assessment scores against the distribution of first-time assessment results across all certificates.
[27]The differences in average effectiveness are not statistically significant in either subject.

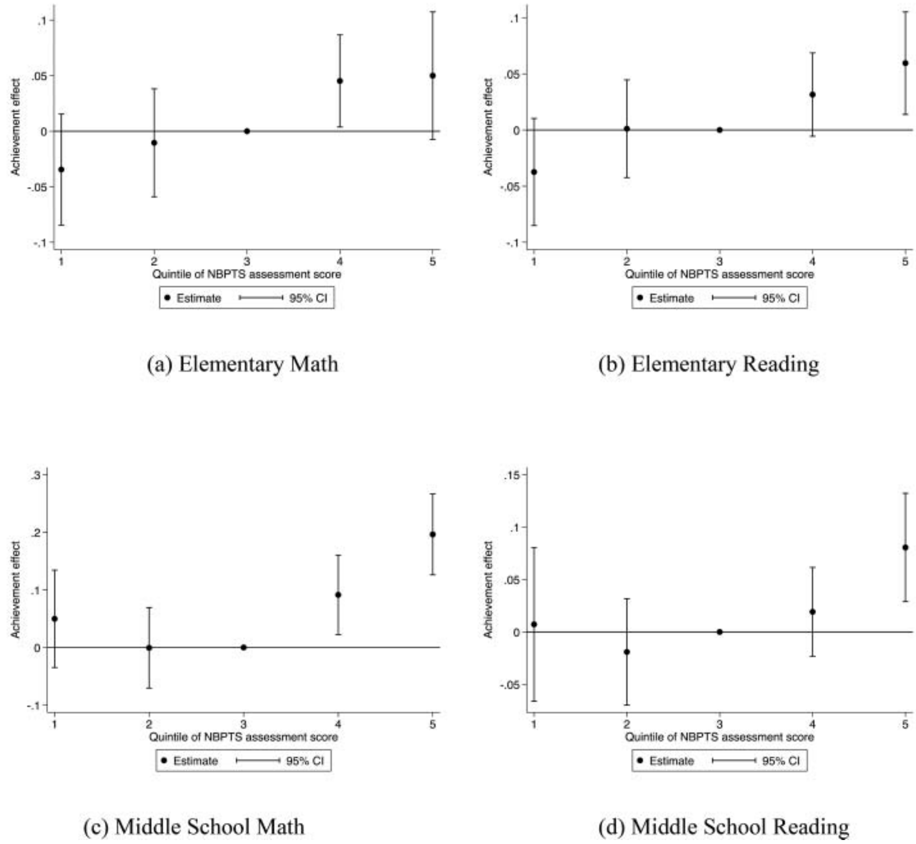

Figure 1. Student achievement effects by NBPTS score quintile.

experience on the sample of NBPTS applicants both with and without the composite candidate assessment score. Without the assessment score, we estimate the variance of teacher effectiveness among National Board applicants is 0.022 in elementary math, 0.015 in elementary school reading, 0.025 in middle school math, and 0.007 in middle school reading. Adding the composite 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 nontraditional measures of preservice teacher quality and find that they explain about 10% of the variation in future teacher effectiveness.

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 generally more effective than non-NBCTS, which is consistent with prior studies of NBCTs in North Carolina and Florida. For elementary math 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. 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.

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). Although 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 & Sass, 2011; Wiswall, 2013).

Although our estimates of the overall effectiveness of NBCTs are broadly consistent with prior work in other states, we present new evidence on the variation of NBCT effects across teachers and students. These results indicate that policymakers ought to be mindful of the variability in effectiveness among teachers who have earned Board certification. We estimate large differences in teacher effectiveness across NBPTS certificate types. Moreover, teachers who possess NBPTS certificates that are uncommon for their teaching assignment generally do not appear more effective than non-NBCTs. We also find that initial performance on the NBPTS assessment provides more information about teacher performance than a teachers' eventual NBPTS status. 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. We also find that the information about teacher effectiveness embedded in the NBPTS assessment is explained by the continuous composite score and that whether the teacher passes or fails does not provide additional information about teacher effectiveness. Each of these findings suggests that policymakers may be justified in considering compensation policies that differentiate between NBCTs, either by targeting performance on the NBPTS assessment rather than the binary certification outcome or varying the incentive based on the certificate earned. Our results indicate that such policies may more discriminately select high-performing teachers.

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. Although 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. It therefore does not appear that increasing the incentive for National Board certification reduces the effectiveness of certified teachers. However, we do not find that NBCTs specialize in teaching less advantaged students, and within-classroom comparisons suggest that NBCTs may be relatively more effective with higher-income students.

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). Although 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 of 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, Glennie, Ladd, & Vigdor, 2008; Dee & Wyckoff, 2013). Further research is needed on the effects of these policies on teacher staffing and their implications for a variety of important student outcomes.

Acknowledgments

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 the data.

Funding

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.

ARTICLE HISTORY

Received 20 May 2015
Revised 18 August 2015
Accepted 18 September 2015

EDITORS

This article was reviewed and accepted under the editorship of Carol McDonald Connor and Spyros Konstantopoulos.

References

Aaronson, D., Barrow, L., & 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., & Wyckoff, J. (2013). *Do first impressions matter? Improvement in early career teacher effectiveness* (NBER Working Paper No. 19096). Cambridge, MA: National Bureau of Economic Research.

Bloom, H.S., Hill, C.J., Black, A.R., & 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., & Staiger, D.O. (2008). *National board certification and teacher effectiveness: Evidence from a random assignment experiment* (NBER Working Paper No. 14608). Cambridge, MA: National Bureau of Economic Research.

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., & 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., & 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., & 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.

Dee, T., & Wyckoff, J. (2013). *Incentives, selection, and teacher performance: Evidence from IMPACT* (Working Paper 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). Denver, CO: National Conference of State Legislatures.

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. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *Journal of Human Resources*, *42*(4), 765–794.

Goldhaber, D., & 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.D., Brewer, D.J., & Anderson, D.J. (1999). A three-way error components analysis of educational productivity. *Education Economics*, *7*(3), 199–208.

Goldhaber, D., & Chaplin, D. D. (2015). Assessing the "Rothstein falsification test": Does it really show teacher value-added models are biased? *Journal of Research on Educational Effectiveness*, *8*, 8–34.

Goldhaber, D., & Hansen, M. (2009). National Board Certification and teachers' career paths: Does NBPTS certification influence how long teachers remain in the profession and where they teach? *Education Finance and Policy, 4*, 229–262.

Goldhaber, D., & 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., Lavery, L., & Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational Researcher*, *44*, 293–307.

Hakel, M. W., Koenig, J. A., & Elliott, S. W. (Eds.). (2008). *Assessing accomplished teaching: Advanced-level certification programs.* Washington, DC: National Academies 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). doi:10.14507/epaa.v13n18.2005
Jackson, C. K. (2012). *Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina* (Working Paper No. 18624). Cambridge, MA: National Bureau of Economic Research.
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.
Johnson, M., Lipscomb, S., & Gill, B. (2015). Sensitivity of teacher value-added estimates to student and peer control variables. *Journal of Research on Educational Effectiveness*, *8*, 60–83.
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., McCaffrey, D. F., Miller, T., & Staiger, D. O (2013). *Have we identified effective teachers?* Seattle, WA: Bill and Melinda Gates Foundation.
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., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation*. NBER Working Paper Series, 14607. Washington, DC: National Bureau of Economic Research.
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.
National Board of Professional Teaching Standards. (2010). *Profiles in excellence: Washington state* (Technical report). Arlington, VA: Author.
National Board of Professional Teaching Standards. (2012). *State profile: Washington* (Technical Report). Arlington, VA: Author.
National Board for Professional Teaching Standards. (2014a). *2014 state rankings by new number of National Board certified teachers*. Arlington, VA: Author.
National Board for Professional Teaching Standards. (2014b). *2014 state rankings by total number of National Board certified teachers*. Arlington, VA: Author.
Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237–257.
Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, *130*, 105–119.
Pearlman, M. (2008). The design architecture of NBPTS certification assessments. In L. Ingvarson & J. Hattie (Eds.), *Assessing teachers for professional certification: The first decade of the National Board for Professional Teaching Standards* (pp. 55–91). Bingley, England: JAI Press.
Protik, A., Walsh, E., Resch, A., Isenberg, E., & Kopa, E. (2013, March). *Does tracking of students bias value-added estimates for teachers?* Paper presented at the Association of Education Finance and Policy Conference, New Orleans, LA.
Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458.
Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, *94*(2), 247–252.

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.
Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics*, *125*(1), 175–214.
Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. (2012). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics*, *72*(2-3), 104–122.
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., & Dillow, S.A. (2013). *Digest of education statistics* 2011 (Technical Report 2014–015). Washington, DC: National Center for Education Statistics.
Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics*, *100*, 61–78.

Appendix

Table A1. Effectiveness of board-certified teachers and board candidates.

	Math			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Elementary Classrooms						
NBCT	0.070***	0.057***	0.079***	0.068***	0.039***	0.050**
	(0.018)	(0.017)	(0.022)	(0.017)	(0.015)	(0.020)
NBCT Candidate	−0.033**	−0.039**	−0.065***	−0.041**	−0.023	−0.045**
	(0.016)	(0.016)	(0.020)	(0.016)	(0.014)	(0.020)
N	742,124	742,124	329,345	742,124	742,124	329,345
Cohort FE	N	Y	Y	N	Y	Y
Apparently random sample	N	N	Y	N	N	Y
Panel B. Middle School Classrooms						
NBCT	0.063**	0.066***	0.060***	0.076***	0.040***	0.046***
	(0.025)	(0.020)	(0.021)	(0.017)	(0.015)	(0.015)
NBCT Candidate	−0.011	−0.014	−0.011	−0.055***	−0.026*	−0.032**
	(0.023)	(0.019)	(0.020)	(0.016)	(0.014)	(0.015)
N	570,533	570,533	570,533	492,800	492,800	492,800
Cohort FE	N	Y	N	N	Y	N
Track FE	N	N	Y	N	N	Y

Notes. Models regress student achievement on indicators 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. "NBCT" indicates the teacher possesses the National Board credentials; "NBCT Candidate" indicates the teacher has applied to NBPTS for board certification. 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.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.