11/29/05

Can Teacher Quality Be Effectively Assessed?
National Board Certification as a Signal of Effective Teaching

Dan Goldhaber
University of Washington and the Urban Institute

Emily Anthony
The Urban Institute

Abstract

In this paper, we describe the results a study assessing the relationship between the certification of teachers by the National Board for Professional Teaching Standards (NBPTS) and elementary-level student achievement. We examine whether NBPTS assesses the most effective applicants, whether certification by NBPTS serves as a signal of teacher quality, and whether completing the NBPTS assessment process serves as catalyst for increasing teacher effectiveness. We find consistent evidence that NBPTS is identifying the more effective teacher applicants and that National Board Certified Teachers are generally more effective than teachers who never applied to the program. The statistical significance and magnitude of the "NBPTS effect," however, differs significantly by grade level and student type. We do not find evidence that the NBPTS certification process itself does anything to increase teacher effectiveness.

(JEL Classification # I20: Education- General)

Acknowledgments

We gratefully acknowledge valuable comments from Dale Ballou; Duncan Chaplin; Drew Gitomer; Tom Kane; Michael Podgursky; Jim Wyckoff; seminar participants at the Economics of Education seminar at Stanford University and at the Peabody College at Vanderbilt University's Department of Leadership, Policy and Organizations; and several anonymous referees. We thank Jim Hunt, former governor of North Carolina and founding chair of the National Board for Professional Teaching Standards (NBPTS), for encouraging North Carolina public officials to participate in the study, and NBPTS for supplying data on National Board teacher candidates. We also thank David Perry, Ashaki Coleman, Karen Callahan, and Gordon Millspaugh for assistance assembling the data set, and Carol Wallace and Michael DeArmond for editorial assistance.

The authors received funding for this study from a grant from the U.S. Department of Education. All opinions expressed in this paper represent those of the authors and not necessarily the institutions with which they are affiliated or the Department of Education. All errors in this paper are solely the responsibility of the authors.

I. Introduction

Education research has failed to reach a consensus over which, if any, readily identifiable teacher characteristics are associated with students' learning gains, and it remains an open question as to whether it is even possible to judge teachers' effectiveness using measures other than direct observations of their teaching. From a policy perspective this is extremely problematic: state-level policymakers lack the knowledge they need to make informed decisions about teacher licensure, and local policymakers lack information that might be useful in hiring teachers and determining compensation. The National Board for Professional Teaching Standards (NBPTS) offers the potential to address some of these issues through the creation of a voluntary certification process whereby teachers who are considered to be highly effective can demonstrate, and gain recognition for, their knowledge and teaching skills.[1]

In both its scope and expense, NBPTS is arguably the most significant national development in teacher policy in the last two decades. Participation in the NBPTS program has grown dramatically over a relatively short period of time: the National Board certified fewer than 200 teachers in 1993–94, but by November, 2005, this figure had climbed to more than 47,500.[2] The program is supported by substantial investments from all levels of government as well as private sources including a substantial number of teachers themselves. NBPTS currently reports federal ($149.1 million) and private ($261 million) funding of $410.1 million.[3] In addition to this, NBPTS receives a fee ($2300 in 2005) for every application, which is paid for either by teachers or, in many cases, by states and localities. Based on an estimated 100,928 applicants (from 1993–94 through 2004–05), this is an additional investment of $227.3 million, putting the total investment in the NBPTS program at upwards of $637 million.[4]

In addition to this direct support of the National Board program, some states and localities offer salary supplements to National Board Certified Teachers (NBCTs).[5] In North Carolina, for instance, NBCTs receive a 12 percent increase in their base pay. In California NBCTs who opt to teach in a "high-priority" school for four years are eligible to receive a $20,000 merit award.[6] While these examples certainly represent the more generous of the direct financial incentives provided to NBCTs, many districts provide other types of incentives (e.g., release time or preparation assistance) that are also costly but more difficult to quantify.

NBPTS offers the potential to address long-standing concerns about teacher quality in at least two ways. First, NBPTS certification may serve as a signal of effectiveness that schools or districts may use to identify skilled teachers for leadership roles or critical teaching positions. The recognition, and often rewards, that accompany the NBPTS credential may also encourage more effective teachers to remain in the teaching profession longer than they otherwise would. Second, the NBPTS certification process itself, which requires applicants to engage in "intense self-reflection and analysis" of their own teaching, may serve as a catalyst for effective teaching by helping to build the human capital of teachers.[7]

While many advocates of NBPTS embrace both the "signal" and "catalyst" hypotheses about the program's benefits, critics argue that there is little evidence to support either view and question the significant investments that have been made in the program (Finn 2003). Because there is surprisingly little quantitative evidence to inform the debate about the program's effects, it certainly remains an open question as to whether policymakers should consider NBPTS as a good signal of teacher quality or a productive way to build the human capital of teachers.

In this paper we hope to inform the debate by presenting the results of the first large-scale study of the program and its relationship to student achievement. Using a unique data set from

North Carolina, we explore both the "signal" and "catalyst" hypotheses about the program's effects. In short, we consistently find that NBPTS certification is successfully identifying effective teachers: NBPTS-certified teachers are more effective than unsuccessful NBPTS applicants and generally more effective than teachers who never applied to the program. The statistical significance and magnitude of the "NBPTS effect," however, differs significantly by grade level and student type. We do not, however, find evidence that the NBPTS certification process itself increases teacher effectiveness.

The paper is laid out as follows: Section II provides some background information on NBPTS, as well as a brief overview of the research literature on the relationship between various teacher characteristics and student outcomes. Section III describes the data and analytic methods we used in the study, and Section IV presents our results. Section V offers policy implications of our findings and some conclusions.

II. Teacher Quality and NBPTS

A growing body of research shows that the quality of the teacher in the classroom is the most important schooling factor predicting student outcomes (e.g., Ferguson 1998; Goldhaber 2002; Goldhaber et al. 1999; Hanushek et al. 1999; Wright et al. 1997). Furthermore, the impact of having a high-quality teacher can be profound. Hanushek (1992), for instance, finds that, all else equal, a student with a very high quality teacher will achieve a learning gain of 1.5 grade level equivalents, while a student with a low-quality teacher achieves a gain of only 0.5 grade level equivalents. Thus, *the quality of a teacher can make the difference of a full year's learning growth.*

While researchers tend to agree that teacher quality is an important determining factor in influencing student outcomes, there is little consensus about the relationship between specific

teacher credentials (e.g., experience and degree level) and characteristics (e.g., age, race and ethnicity), and teacher effectiveness.[8] Certain teacher characteristics—a subject-specific master's degree, for example—seem to matter in certain contexts (e.g., high school math), but not in others. More to the point, only a small percentage of what makes teachers successful in the classroom is associated with characteristics such as degree and experience levels, and certification status (Goldhaber 2002; Goldhaber and Brewer 2000; Hanushek 1986, 1997). In other words, teachers clearly matter, but *teacher quality* is not strongly related to *observed teacher characteristics.* There is a seeming contradiction between the fact that teachers have a large impact on student achievement but specific teacher attributes are not consistently found to directly impact student achievement. This may be credited to the fact that the attributes that actually make teachers successful in the classroom (e.g., enthusiasm and ability to convey knowledge) are not strongly related to the teacher attributes typically measured in education productivity studies.[9]

NBPTS was founded on the idea that the attributes that make experienced teachers effective can, in fact, be identified and evaluated. A casual look at the program suggests that demonstrated mastery of NBPTS standards represents a significantly higher hurdle than does standard state-level licensure. From 1999 through 2002, only about 50 percent of first-time NBPTS applicants became certified[10] while about 90 percent of teachers typically pass states' licensure exams (e.g. Praxis I or II) on the first try.[11] Furthermore, research suggests there is a strong correlation between an NBPTS applicant's performance on standardized tests and both the probabilities of application and, given application, NBPTS certification (Goldhaber et al. 2004). Given that numerous studies find a positive connection between teacher performance on measures of academic proficiency and student outcomes, this provides some face validity to the

notion that the program is identifying better teachers (Ferguson and Ladd 1996; Goldhaber 2002; Greenwald et al. 1996). Of course, these facts are only suggestive. A real test of the value of NBPTS as an indicator and/or catalyst for effective teaching requires that we look at a direct measure of teacher effectiveness: student outcomes. As NBPTS skeptics often note, to date, such evidence is lacking (Ballou and Podgursky 1998; Podgursky 2001).

Given educational resource constraints and the size of the local, state, and national investment in NBPTS, policymakers have reason to be concerned about whether NBPTS certification is in fact an effective indicator of teacher quality or has the potential to increase teacher effectiveness by building human capital. We are aware of only a limited number of studies (Bond et al. 2000; Cavalluzzo 2004; Stone 2002; Vandevoort et al. 2004) that attempt to link NBPTS certification status directly to student outcomes. Of these, only Bond et al. and Vandevoort et al. have been published in a peer-reviewed journal, and each suffers from serious data shortcomings: Bond et al. is based on a sample of 31 NBCTs, while the Vandervoort et al. study includes only 35 NBCTs, and neither study includes statistical adjustments for differences in student demographics (Vandevoort et al. measures student gains in performance while Bond et al. does not). Thus, the available literature on NBPTS has been striking in its absence of rigorous quantitative studies that policymakers might use to judge the relative costs and benefits of the NBPTS program, despite its potential to identify teaching skills that may relate to student learning. As we describe below, our study sample includes longitudinal information on a significantly larger sample of students and teachers, which permits the estimation of more sophisticated statistical models that account for potential biases associated with the possibility that students are nonrandomly assigned to their teachers.

III. Analytic Approach and Data

A. Analytic Approach

Our metric for measuring the effectiveness of NBCTs utilizes student performance on standardized tests administered as part of the North Carolina accountability system.[12] We begin by estimating a basic educational production function of the following form:[13]

$$A_{ijst} - A_{ijs(t-1)} = \alpha X_{it} + \beta NBPTS_{ijt} + \gamma S_{ist} + \nu_{ijst} \quad (1)$$

The subscripts *i*, *j*, *s*, and *t* denote individual student, teacher, school/district/community, and time, respectively. The left-hand side of the equation ($A_{ijt} - A_{ij(t-1)}$) is the growth in student test score from time (*t*-1) to time *t*. *X* (from here forward we suppress the subscripts for simplicity) is a vector of individual characteristics including student's race, gender, learning disability, free or reduced-price lunch status,[14] English proficiency status, grade, and year. NBPTS is a vector of characteristics defining the National Board status of student *i*'s teacher, and *S* is a vector of other teacher, school, and community control variables including the teacher's race/ethnicity, gender, age, license basis and status, degree level, dichotomous variables for various years of teaching experience, standardized test scores, school size (number of students), school student-to-teacher ratio, fraction of minority students at school, fraction of free or reduced-price lunch at school, district size (number of students), the expenditure per pupil in the district, district type (urban, suburban, or rural), the percent of education expenditure spent on instruction, the starting salary of teachers with a bachelor's degree in that district, the percent of people with a bachelor's degree in the community, and the median housing value in the community.[15]

We investigate several hypotheses using the above model. First, by comparing NBCTs to non-NBCTs (both noncertified applicants and non-applicants), we examine whether or not NBPTS certification is a good signal of teacher quality in the teacher labor market as a whole.

From a policy perspective, this comparison is particularly relevant given that the federal government has identified NBPTS certification as a tool that states can use to meet the teacher quality provisions of the No Child Left Behind Act.[16] Second, we examine whether NBPTS certification provides any information about teacher effectiveness that goes beyond what is already provided by teacher licensure tests (in other words, are relatively cheap licensure tests just as good as NBPTS at identifying effective teachers?).

Next we examine the question of whether NBPTS is identifying the most effective candidates among those who apply for the program by comparing successful and failed applicants. NBPTS only makes judgments about those who apply to the program, and evidence suggests that the NBPTS applicant pool is quite different from the teacher workforce as a whole (Goldhaber et al. 2004). Finally, we investigate the question of whether or not the NBPTS certification process itself is a catalyst for building human capital by comparing teacher effectiveness before and after completing the NBPTS certification process.[17]

Given our concern that NBCTs may have differential impacts on different types of students, and that our findings may be confounded by nonrandom sorting of students across teachers (Clotfelter, Ladd, and Vigdor 2003), we also estimate models for different subgroups of students and specify models that include school and student fixed-effects.

B. Data

The primary source of data for this study is teacher- and student-level administrative records from North Carolina's Department of Public Instruction (NCDPI) for school years 1996–97 through 1998–99 (in addition, we have information on whether teachers in our sample applied to the program or became NBPTS-certified through 1999—2000).[18] North Carolina is an ideal state for studying the effects of NBPTS certification due to the large numbers of NBCTs in the state,

and because the state accountability system requires yearly testing of students using aligned tests to track progress over several years.[19] Furthermore, it is possible with these data to link teacher and student records (at the elementary level) and to track both over time.[20]

The NCDPI teacher records include variables such as teacher's race/ethnicity, gender, age, license basis and status, degree level, years of teaching experience, and a measure of teacher academic proficiency—that is, their performance on one or more standardized tests including one or more of the following: the Praxis generalist test (Praxis I), Praxis subject tests (Praxis II), the National Teacher Exam (NTE), and, in some cases, teachers' SAT and GRE scores.[21] We convert the various test scores into Z-scores in order to place them on a common metric, and experiment with using various test's Z-scores as our measure of teacher academic proficiency.[22] We use the average of teachers' Praxis I (if present on teacher record) and Praxis II Z-scores (henceforth referred to as "teacher Z-score") as controls for teacher quality.[23]

Teacher records from NCDPI are then matched to information obtained from the Educational Testing Service (ETS), which maintains NBPTS certification information for NBPTS teacher applicants. The ETS teacher records include the year in which teachers applied, the NBPTS certification area to which they applied, and if the teachers were ultimately successful in the process.

In linking the NBPTS records to state teacher records, it was necessary to decide in which year it is appropriate to classify a teacher as being NBPTS-certified, because the application and certification process generally happens over the course of two school years.[24] Based on the NBPTS application and certification timeline (shown in **figure 1** of appendix A), we opted to classify teachers' certification status as the school year in which they completed the NBPTS requirements rather than the school year in which the results are announced, because the bulk of

the work for becoming NBPTS-certified (completing the application) occurs in the school year prior to the one in which certifications are announced.[25]

The student records maintained by NCDPI contain student background information such as student's race/ethnicity, gender, learning disability, free or reduced-price lunch status (available from the state in school year 1998–99 only), English proficiency status, grade and year, and test results for grades 3 through 10. The tests are designed to measure subject objectives defined in the *North Carolina Standard Course of Study*, and are used by the NCDPI's Accountability Department as part of the "ABC" education reform program to determine performance and growth/gain goals and ratings for all schools in the state. All tests are vertically aligned, allowing us to determine individual student achievement growth in addition to school growth performance by subtracting the previous year's end-of-grade test from the subsequent year's end-of-grade test in that subject.[26]

We opted to restrict our study to elementary students in the 3rd, 4th, and 5th grades because elementary-level students are most likely to have only one teacher per grade, thus enabling us to link students' records to their teachers. Our linkage of students and teachers was very successful, yielding pre- and post-test scores for a large number of students.[27] In **table B.1** in appendix B, we report the number of student and teacher observations in each year, the number of these records that we were able to match together and over time, and the number for which we have both a valid end-of-year test score and a pre-test score (either the beginning of the year in the case of the 3rd grade, or the end of the previous year for the 4th and 5th grades).[28] Of the NBCTs in our teacher observations, almost all of them have a NBPTS Generalist Certificate.

Overall, we were able to match 771,537 of the 889,655 student observations with their teachers (for three grades from the 1996–97 school year to the 1998–99 school year), which is about an 80 percent match rate. Of these, we matched 609,160 student observations with 32,399 teacher observations that included valid scores for the reading pre- and end-of-year test, and 611,517 student observations with 32,448 teacher observations that included this same information for math. Appendix **table B.2** reports the number of teachers in each year who fall into different NBPTS certification categories: future, current, and past applicant, and future, current, new, and past NBCT. Because the program is growing rapidly and we have an extra year of information about teachers (the 1999–00 school year), in every year there are many more future applicants and future NBCTs in our sample than past applicants or past NBCTs.[29]

Table 1 presents student means by NBPTS teacher certification status for selected student and teacher variables. Roughly 9,000 unique students in our sample of 390,449 unique students have a teacher who taught them while she was going through the NBPTS assessment process. Approximately 6,000 students in our sample have a teacher who was successfully NBPTS-certified by the time that teacher taught them. Although students with NBCTs have, on average, higher end-of-year test scores in both math and reading, they also tend to have higher initial pre-test scores. Still, the growth in both reading (6.18 points) and math (10.21 points) performance for students who have NBCTs was slightly higher (the difference was statistically significant at the 1 percent level) than the growth for both those who have nonapplicant teachers (5.69 and 9.75 for reading and math, respectively) and those who have teachers who were unsuccessful applicants (5.83 and 9.14 for reading and math, respectively, and again the difference is statistically significant at the 1 percent level). These differences, nonetheless, are relatively small; the largest differential is in math between certified and noncertified applicants,

at just over a point on the exam or roughly 14 percent of a standard deviation in the growth in math scores.[30]

Some of these differences in test scores may be explained by factors other than the certification status of teachers. For example, NBCTs tend to be teaching in more affluent, well-educated school districts, and they are teaching in schools that have been judged by the state to be high-performing and that have fewer disadvantaged students. Furthermore, NBCTs themselves differ from non-NBCTs in that they are likely to have performed far better on any of the teacher licensure exams. In the next section, we explore whether any of these factors explain the differential in students' average gain scores between NBCTs and non-NBCTs.

IV. Results

Prior to focusing on the NBPTS variables, it is worth discussing some of the other individual student and school findings.[31] On both reading and math tests, students who are black, female, participants in the free and reduced-price lunch program, and/or have learning disabilities do worse than students who are white, male, nonparticipants in the free/reduced-price lunch program, and/or do not have learning disabilities.[32] Variables identifying years of teaching experience and having a "full" teaching license from the state (as opposed to a provisional or temporary license, or one that only meets the state's initial teaching license requirements) are generally positive and significant. The signs and statistical significance levels for many of the other schooling variables, however, are sensitive to model specification and often counter-intuitive, which is consistent with much of the educational productivity literature.[33] For example, in some model specifications we find that a teacher having an advanced degree is detrimental to student achievement and that students benefit from being in either larger dasses or larger schools.

Table 2 shows coefficient estimates for key teacher and NBPTS variables in reading and math achievement models. Columns 1 through 6 show the results for various specifications of the growth in the reading achievement model, and columns 7 through 12 show various specifications for the growth in the math achievement model. Prior to discussing these results, it is worth noting that the results discussed below are quite insensitive to the elimination of students with learning disabilities or who are limited English dependent.[34] Further, the estimation of models where the post-test score is regressed on a lagged test score along with the other specified controls generates NBPTS findings that are not materially different from those that are reported.

A. NBPTS Certification as a Signal of Effective Teaching

We begin by assessing how NBCTs compare to non-NBCTs, which is the comparison of interest for policymakers who may wish to use the NBPTS credential as a signal of teacher quality. Recall that we can compare NBCTs both prior to their receiving certification and after they have been identified by NBPTS as having mastered the standards for "what accomplished teachers should know and be able to do." Thus, in the first specification of the reading (column 1 of **table 2**) and math (column 7 of **table 2**) models we include two NBPTS variables: whether a teacher in our sample is an NBCT ("current NBCT"), and whether a teacher who is not currently certified becomes an NBCT at some point in the future, up through the 1999–2000 school year ("future NBCT").[35] The excluded comparison group in this specification (non-NBCTs) includes those teachers who are either non-applicants or who apply to the program but fail to achieve certification.

The magnitudes of the future NBCT coefficients suggest that student gains produced by NBCTs exceed those of noncertified applicants by about 4 percent of a standard deviation in

reading and 5 percent of a standard deviation in math (based on a standard deviation of 9.94 on the end-of-year reading tests and 12.34 on the end-of-year math tests).[36] These effects sizes are of the same order of magnitude as those found for math teachers having a bachelor's degree in their subject area (Goldhaber and Brewer 1997). The findings for current NBCTs are smaller but still positive, and in the case of the reading model, statistically significant (we discuss possible reasons for differences in the estimated coefficients of future and current NBCTs below).

Our finding that NBCTs appear more effective than non-NBCTs is perhaps not surprising, given the aforementioned research linking measures of teachers' academic proficiency to students' achievement, and previous research showing a strong correlation between teachers' performance on licensure tests and both the likelihood of application to NBPTS and eventual certification of teacher applicants (Goldhaber et al. 2004). It does, however, raise the question of whether NBPTS certification conveys information about teacher quality above and beyond that which is learned from teachers' licensure test performance. This is an important public policy question, since states and localities might simply use licensure performance in place of NBPTS certification were it to provide as much information about teacher quality. This would, of course, allow for considerable savings: the cost of the NBPTS assessment alone is $2,300, and many states and localities also provide salary supplements for their certified teachers.

To address this issue, we report specifications of the reading (column 2 of **table 2**) and math (column 8 of **table 2**) achievement models that include a control for teachers' licensure performance (their Z-score). Surprisingly, in both models the magnitude of the coefficient estimates on the NBPTS variables diminish only slightly with the addition of this teacher-quality

control.[37] This suggests that NBPTS certification does in fact convey information about teacher quality above and beyond what can be learned from performance on teacher licensure tests alone.

One of the surprising results from these models is that future NBCTs appear to be far more effective prior to receiving their certification than after they have received it (based on the difference in magnitudes of the coefficients of future and current NBCT). A finding that the coefficient on current NBCT is not larger than future NBCT would suggest that NBPTS certification does not add to teachers' human capital (an issue we explore in greater depth below), but our findings actually suggest that teachers destined for certification are more effective *before* they are recognized by NBPTS. One possibility for this seemingly strange result is that the time intensity of the NBPTS assessment process makes NBCTs less effective in the year in which they receive certification, because they are allocating a significant amount of time to completing the assessment that would otherwise be allocated toward teaching.

We test the hypothesis that teacher effectiveness may be influenced by the time taken to complete the NBPTS assessment by estimating specifications of the model that include four NBPTS variables: future NBCT (defined above), whether a teacher is an applicant to NBPTS in year *t* ("current applicant"), whether a teacher is certified in year *t* ("new NBCT"), and whether a teacher was certified in a year prior to year *t* ("past NBCT"). The omitted comparison group in this model specification is non-applicant teachers. This model specification is reported in column 3 of **table 2** for reading and column 9 of **table 2** for math.

In addition to allowing us to test the time-allocation hypothesis (an issue explored in greater detail in the next subsection), this specification has the added benefit of providing a measure of whether NBPTS is identifying the more-effective teachers among those that actually apply to the program. Recall that NBPTS is making judgments only about those teachers who

have applied for certification, so a comparison between current NBCTs and noncertified teachers ignores the possibility that the NBPTS applicant pool might be very different from the teacher workforce as a whole. Thus, to judge the effectiveness of the NBPTS assessment process, we wish to compare NBCTs to unsuccessful NBPTS applicants, a comparison that is identified in this model specification by the coefficient of new NBCT.

The positive significant coefficient for new NBCT, in both reading and math models, indicates that teachers who are successful in their attempt to attain certification are more effective than those who are unsuccessful applicants, providing evidence that NBPTS is in fact identifying the more effective teachers of those they actually evaluate. The magnitude of the coefficients suggests that students of NBCTs would be expected to achieve growth exceeding that of students of unsuccessful applicants by about 5 percent of a standard deviation in reading and 9 percent of a standard deviation in math.

The primary reason for the differential between certified and uncertified teacher applicants is that teachers who apply to the program but are unsuccessful in their attempt at certification are actually less effective than non-applicant teachers (this effect is identified by the coefficient on current applicant). The total effect on students of having an NBCT in the year in which they apply to the program—the sum of the coefficients of current applicant and new NBCT—is not statistically different than zero, implying that NBCTs are no more or less effective than non-NBCTs when they are going through the NBPTS assessment process. These findings provide some evidence that the time required to complete the NBPTS assessment does have at least a short-term negative impact on teacher effectiveness.

We might also expect that teachers who have achieved NBPTS certification would be more effective than non-applicants in the years after completing the process (as they were found

to be pre-assessment); a measure of this is the coefficient on past NBCT. However, neither our reading nor math results support this hypothesis: the magnitude of the past NBCT coefficient is smaller (and not significant) than the future NBCT coefficient in the reading model, and in the math model it is actually negative (but not significant).[38] While we interpret these findings with caution because our data set has very few teachers who fall into the past NBCT category,[39] the seemingly strange result that NBCTs are less effective after they receive certification than before merits further investigation.

There are at least two possible explanations for these anomalous findings that imply a bias in the estimates of the NBPTS variables. One is that teachers may be encouraged by principals or other district officials to apply to NBPTS based on temporary upward blips in their performance. If this tends to be the case, then the findings may simply represent a phenomenon akin to a reverse "Ashenfelter dip" (Ashenfelter and Card 1985) where the future NBCT variable is biased upward and the smaller coefficient on past NBCT simply reflects teachers reverting to their typical level of performance.[40] A second explanation is that NBCTs tend to be assigned to different types of students. If, for instance, they are systematically assigned to higher-achieving students (those with higher pre-test scores) and there is mean reversion in student test performance, then the estimated coefficient of past NBCT would have a downward bias. We explore these possibilities in the following subsections of the paper.

B. Specification Checks: Does the NBPTS Assessment Add to Teacher Human Capital?

One way to measure whether going through the NBPTS assessment process makes a teacher more effective is to compare the coefficients of past and future NBCT. As we discussed above (in reference to columns 3 and 9 of **table 2**), the results of this comparison do not support this idea, since NBCTs appear to be more effective before they are certified than after. Models that

include teacher fixed-effects tend to confirm these findings.[41] In a teacher fixed-effects math model the estimated coefficient of past NBCT (-.74) is statistically significant and negative, suggesting that NBCTs are not more effective teaching math than they were prior to completing the certification process (none of the coefficients were statistically significant in a teacher fixed-effects reading model). Again, these results should be interpreted very cautiously, since the impact of teachers' NBPTS status is identified by variation over time in teachers' NBPTS status and our three-year panel, of course, contains little variation.

We further explore the question of whether the NBPTS process itself adds to a teacher's human capital by estimating models that test whether *any* applicants—either successful or unsuccessful—benefit from going through the assessment process. Specifically, we estimate a model specification (reported in column 4 of **table 2** for reading, and column 10 of **table 2** for math) that includes three NBPTS *application* variables: future applicant, current applicant, and past applicant (non-applicants are the excluded comparison group). As was the case with certification, we might expect that if teachers accumulate human capital as a consequence of the assessment process they would be more effective post- than pre-assessment, whether they are successful or not in attaining certification. And, as we discussed above, it would not be surprising to find applicants to be less effective than non-applicants because of the time they are allocating to complete the NBPTS assessment.

The pattern of results provides strong support for the hypothesis that the time required to complete the NBPTS assessment negatively impacts teacher effectiveness. In both math and reading, teacher applicants are significantly less effective in the year of application than they are in either the pre- or post-application year. The test of human capital effects among applicants also provides no evidence that completing the NBPTS assessment increases teacher

effectiveness. The coefficients on past applicant are not greater in magnitude than the coefficients on future applicant in either the reading or math models, but again this is based on a relatively small sample of past applicants.[42]

As discussed in the previous subsection, there are at least two potential sources of bias in our estimates of the NBPTS variables that could explain why we do not find stronger human-capital effects associated with the NBPTS process. One is that teachers might be encouraged by principals or other district officials to apply to NBPTS based on a temporary upward blip in their performance, and thus their post-certification performance reflects a return to their baseline. To test for this, we estimated a model that included separate indicator variables for each of the years prior to application to determine whether the year immediately prior to application looks to be significantly different from the pre-application trend in teacher effectiveness. While teachers do appear to be more effective in the year prior to application than they are during or post-application, there is only limited evidence that they are more effective in the year immediately prior to application. In reading and math, teachers appear to be slightly more effective in the year prior to application than they were two years prior to application (the difference in magnitude between the coefficients on one-year prior to application and two-years prior to application is significant at the 10 percent confidence level in reading and at the 5 percent level in math), but there was no significant difference found between the year prior to application and three years prior to application. Thus, it does not appear that the year prior to application represents an unusual upward deviation from the long-term trend in teacher effectiveness.[43]

We might also imagine that differences in teacher performance before and after completing the NBPTS assessment process simply reflect differences in the effectiveness of various teacher cohorts. For instance, all "past applicants" are teachers who went through the

NBPTS assessment process in 1998–99 or an earlier year, while all "future applicants" are teachers who went through the NBPTS assessment process in 1997-98 or a later year. We attempt to distinguish between cohort and human-capital effects by estimating models that interact the application and year variables to determine whether applicants to the program appeared to be more effective teachers in some years than others. While there are some statistically significant interaction terms, these models do not appear to show any systematic relationship between the year of application and the measures of teacher effectiveness.

A final potential source of bias is nonrandom teacher assignment, and specifically the possibility that NBPTS status affects the assignment of teachers. We explore this possibility in the models presented below.

C. School and Student Fixed-Effects Model Specifications

A significant amount of empirical evidence shows that teachers are not randomly distributed across students. Nonwhite, poor, and low-performing students are more likely to be taught by less-qualified teachers, as measured by experience and degree levels, licensure status, licensure exam performance, and college selectivity (Lankford et al. 2002). This sorting pattern occurs both within and between districts, and the movement of experienced teachers between schools and districts tends to worsen inequities because more highly qualified teachers are found more likely to leave poor urban schools to teach in higher-performing, more affluent schools. Research by Levinson (1988) and Hanushek et al. (2004) finds that students' socioeconomic status and achievement play an important role in explaining the schools teachers choose as employers when they move from one school or district to another, and Goldhaber et al. (forthcoming) and Clotfelter et al. (2003) suggest that these nonrandom sorting patterns may be even more pronounced for NBCTs.

If teachers and teacher quality are nonrandomly distributed across students and student characteristics, as the evidence strongly suggests, and statistical models do not fully account for the student characteristics affecting achievement, then the estimated effects of observed teacher characteristics are likely biased. In fact, evidence suggests that the estimated impact of teacher credentials tends to be overstated due to nonrandom student and teacher assignment (Clotfelter et al. 2003).

To take account of the nonrandom distribution of teachers across schools, we estimate specifications of the reading and math achievement models that include school fixed effects. The estimated teacher-variable coefficients in these models are identified based on within-school variation in teacher characteristics. Column 5 of **table 2** shows the results of these models for reading growth, and Column 11 of **table 2** shows the results for math growth. The inclusion of these school effects makes little difference in the estimates of NBCT effects. In both reading and math, we still find that teachers who will eventually be certified (future NBCT) are more effective prior to certification, and certified teachers are more effective than noncertified applicants (the coefficient of new NBCT) but no more effective than non-applicants (the sum of current applicant and new NBCT). There is no case where the magnitude of an estimated coefficient is statistically different from the base models, suggesting that most of the findings are not influenced by a nonrandom distribution of NBCTs across schools.[44]

These models account for nonrandom sorting of teachers across schools, but they do not account for the possibility that teachers may be nonrandomly sorted across students within schools. One could easily imagine this would occur due to seniority-based assignments, or to the pressures from parents to assign their students to particular teachers. We account for this within-school sorting by estimating specifications of our models that include student fixed effects. In

these models, the impacts of teacher characteristics are identified by variation over time in the characteristics of the teachers to which students are assigned. Columns 6 and 12 of **table 2** present the student fixed-effects specifications for reading and math, respectively.[45] The results of these models continue to show that NBCTs are more effective before they are recognized by NBPTS (based on the coefficient on future NBCT), however, we do find some notable differences in the estimated effects of newly and previously certified NBCTs. Specifically, in reading, newly certified teachers do not appear to be more effective than noncertified applicants (based on the coefficient on new NBCT), and previously certified NBCTs are no more effective than non-applicants. In math, we now find that previously certified NBCTs are actually less effective than are non-applicants. While these results must be interpreted with some caution given that they are based on a three-year panel (so there is relatively little time-series variation by which to identify these effects) and there are relatively few previously certified NBCTs in the data, they are nevertheless puzzling. One possible explanation for the findings on new and past NBCTs is mean-reversion of scores among high-achieving students. If mean-reversion exists and NBCTs are systematically assigned to students with high pre-test scores, as appears to be the case (Clotfelter et al. 2003; Goldhaber et al. forthcoming), then mean-reversion would imply a negative bias in the estimates of the effects of NBCTs.

D. Results by Student Subgroup and Grade Level

There are several reasons why the effects of new NBCTs and future NBCTs on students might vary by student subgroup or grade level. Empirical evidence dating from the "Coleman Report" (Coleman et al. 1966) tends to find that teacher quality has a larger impact on lower-achieving students than on those who are higher-achieving, and this finding has recently been confirmed by empirical work drawing on an experiment where students are randomly assigned to teachers

(Nye et al. 2004). Furthermore, NBPTS is thought to value particular constructivist approaches to teaching (Wilcox 1999; Ballou 2003) that may be more or less effective when employed with different types of students or with students of varying ages or academic achievement levels. Whether the impact of NBCTs varies by student subgroup or grade level is also an important public policy concern: educational administrators need this information in order to allocate NBCTs effectively across students and grades. For all these reasons, we present results that are broken out by student subgroup (free and reduced-price lunch status and race) (**table 3**) and grade level (**table 4**).[46]

Our findings for various student subgroups are consistent with previous findings that teacher quality has a larger impact on poor students than on higher-income students (Coleman 1990). The magnitude of the effect of having a new NBCT or a past NBCT is significantly larger in reading for students who are receiving free or reduced-price lunch than those who are not (comparing columns 1 and 3 of **table 3**), and is significantly larger for having a new NBCT or a future NBCT in math for students who are receiving free or reduced-price lunch than those who are not (comparing columns 2 and 4 of **table 3**).

There are also some notable differences in findings for students of different races or ethnicities. Future NBCT is positive and significant across all student race groups, but the effects are largest for students in the "other" category, where the point estimates for the reading and math models (.83 in reading and .96 in math) are statistically different from the point estimates for white students (.34 in reading and .65 in math) and black students (.31 in reading and .55 in math) (see **table 3**).[47] Though not reported here, we also estimated student and school fixed-effects specifications of the above models. The pattern of results (the direction of the estimated effects) for white students is generally similar, though the levels of statistical significance

change, whereas there are no consistent results for the various categories of minority students, which is not surprising given the small numbers of minority students who have NBCTs.

When examining the grade-level models, the overall patterns of results tend to be consistent with our findings above: we see large positive effects for future NBCTs, some positive NBPTS identification effects but little or no difference between NBCTs and non-applicants in the year of application, and divergent results in math and reading for past NBCTs. We also observe important differences in the estimated effects by grade level. In general, the largest significant positive NBPTS effects are found in the 3rd grade. For example, while the future NBCT coefficient is positive and statistically significant for all grades and both subjects, the magnitude of this coefficient is significantly larger in the 3rd grade than either the 4th or 5th grades (comparing across columns in **table 4**). The new NBCT coefficient is significant and positive for both reading and math in the 3rd grade, but outside of the 3rd grade it is only positive and significant for 4th grade math.[48] The differential impacts of new NBCTs and future NBCTs by grade level have important policy implications: these results at least suggest that greater benefits are provided to students if NBCTs are assigned to teach the earlier elementary grades.

V. Public Policy Implications and Conclusions

The significant interest, investment, and growth in NBPTS certification represents an important effort to try to professionalize teaching, bring effective educational practices to the classroom, and ultimately enhance student learning. But, until now there has been little quantitative research assessing whether NBPTS certification is actually associated with student outcomes. In this paper we consider NBPTS certification both as an indicator of and as a catalyst for teacher effectiveness. Our findings appear to confirm that there is value to the NBPTS assessment

process in distinguishing between more- and less-effective teachers: teachers who are certified by NBPTS tend to be more effective than unsuccessful applicants to the program. These results are robust to a number of different model specifications, including those that account for the potential nonrandom matching of students to schools or classrooms. While we consistently find that teachers who will eventually be NBPTS-certified are more effective, there are mixed findings about their effectiveness post-certification—making judgments about the "signal" hypothesis less certain. As for the hypothesis that NBPTS certification acts as a catalyst for effective teaching, by building teachers' human capital, we find no evidence to support the notion that completing the NBPTS assessment process increases teacher effectiveness.

Our findings that NBCTs appear to be no more effective (and in some cases less effective) post-certification than they were pre-certification are rather puzzling. As we discussed above, these findings may simply be an artifact of the relatively small group of teachers in our sample who were teaching (post-certification) having already completed the NBPTS process. They may also reflect one or more of the potential sources of bias we discussed (though the checks we were able to perform with the available data appear to rule these out). One possible explanation for why NBCTs appear to be less effective after becoming certified is that teachers who are certified may, post-certification, devote a greater share of their time and effort to endeavors outside of their own classes. Although an investigation of this issue is beyond the scope of this study, the notion that NBCTs might be taking on new or different responsibilities is consistent with the National Board's original conception of the roles that NBCTs might play in schools—for instance, serving as mentor teachers or curriculum specialists (http://www.nbpts.org/about/prop5). It is also consistent with findings from surveys of NBCTs,

which suggest that they do take on more schoolwide roles upon becoming certified (http://www.nbpts.org/pdf/leadership_survey.pdf).

While the main conclusion that people may take from this paper might be, "National Board successfully identifies effective teachers," policymakers would do well to look more closely. Ultimately, whether or not the investment of public dollars in NBPTS makes sense depends on more nuanced findings that suggest its benefits differ by grade level and student type. We find, for example, that schools with NBCTs receive substantially more educational benefits from having their NBCTs teach low-income students in earlier grades.[49] It may also depend on the effects of applying to the program since we find that unsuccessful current applicants are actually less-effective teachers in the year in which they apply to NBPTS. Schools with many unsuccessful applicants, or those with successful applicants who leave after obtaining certification, may actually be worse off for having had their teachers apply to the program (recall that NBCTs are no more, or less, effective than non-applicants in the year of application, and unsuccessful applicants are actually less effective than non-applicants while they are completing the assessment process). Considering this, the total impact of the NBPTS program on schools will depend on the assignment and retention of NBCTs as well as the number of successful and unsuccessful applicants in a given school. These more subtle findings suggest that understanding the impact NBPTS has on students requires not just an assessment of its worth as an indicator of quality or a catalyst for building human capital. It also requires an understanding of how NBPTS affects teachers' career paths: Who do NBCTs teach? How long do they teach? Do the roles of NBCTs extend beyond the classrooms in which they are assigned to teach? All of these are important questions for further empirical study.

References

Ashenfelter O., and D. Card, "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs." *Review of Economics and Statistics* 67:4 (1985): 648-660.

Ballou, D., "Certifying Accomplished Teachers: A Critical Look at the National Board for Professional Teaching Standards." *Peabody Journal of Education* 78:4 (2003), 201-219.

Ballou, D. and M. Podgursky, "The case against teacher certification." *The Public Interest* 132 (1998), 17–29.

Bond, L., T. Smith, W. Baker, and J. Hattie, *The Certification System of the National Board for Professional Teaching Standards: A Construct and Consequential Validity Study*. University of North Carolina at Greensboro, Center for Educational Research and Evaluation (2000).

Card, D., and A. B. Krueger, "Labor Market Effects of School Quality: Theory and Evidence," in G. Burtless (ed.), Does Money Matter? The Link Between Schools, Student Achievement and Adult Success (Washington, D.C.: Brookings Institution, 1996), 97-140.

Cavalluzo, L C., (2004) "Is National Board Certification An Effective Signal of Teacher Quality?" CNA Corporation. (available online at http://www.cna.org/documents/CavaluzzoStudy.pdf)

Clotfelter C.T., H. F. Ladd, and J.L. Vigdor, "Teacher sorting, teacher shopping, and the assessment of teacher effectiveness." Paper presented at the American Association of Public Policy and Management (November 2003).

Coleman, J.S., *Equality and Achievement in Education* (Boulder, Co: Westview Press, 1990).

Coleman, J.S., E. Campbell, C. Hobson, J. McPartland, A. Mood, F. Weinfeld, and R. York, *Equality of Educational Opportunity* (Washington, DC: U.S. Government Printing Office, 1966).

Ferguson, R., "Teachers' perceptions and expectations and the Black-White test score gap." In C. Jencks and M. Phillips (Eds.), *The Black-White Test Score Gap* (Washington, DC: Brookings Institution Press, 1998), 273–317.

Ferguson, R. and H. Ladd, "How and why money matters: An analysis of Alabama schools." In Helen Ladd (Ed.), *Holding Schools Accountable* (Washington, DC: Brookings Institute Press, 1996).

Finn, C. E. Jr., "High Hurdles." *Education Next* 3:2 (2003), 62-67.

Goldhaber, D., "The Mystery of Good Teaching: Surveying the Evidence on Student Achievement and Teachers' Characteristics." *Education Next* 2:1 (2002), 50-55.

Goldhaber, D. and D.J. Brewer, "Evaluating the effect of teacher degree level on educational performance." In J. William Fowler (Ed.), *Developments in School Finance 1996* (Washington, DC: National Center for Education Statistics, 1997), 197 - 210.

——— "Does teacher certification matter? High school teacher certification status and student achievement." *Educational Evaluation and Policy Analysis* 22 (2000), 129-145.

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

Goldhaber, D., H.J. Choi, and L. Cramer, "A Descriptive Analysis of the Distribution of NBPTS Certified Teachers in North Carolina." *Economics of Education Review* (Forthcoming).

Goldhaber, D., D. Perry, and E. Anthony, "NBPTS Certification: Who Applies and What Factors are Associated with Success?" *Educational Evaluation and Policy Analysis,* 26:4 (2004), 259-280.

Greenwald, R., L. Hedges, and R. Laine, "The effect of school resources on student achievement." *Review of Education Research* 66:3 (1996), 361-396.

Hanushek, E. A., "The economics of schooling: Production and efficiency in public schools." *Journal of Economic Literature 24*:3 (1986), 1141-78.

——— "The trade-off between child quantity and quality." *Journal of Political Economy* 100:1 (1992), 84-117.

——— "Assessing the effects of school resources on student performance: An update." *Educational Evaluation and Policy Analysis* 19:2 (1997), 141-64. (EJ 550 073).

Hanushek, E. A., J. F. Kain, and S.G. Rivkin, *Do higher salaries buy better teachers?* National Bureau of Economic Research Working Paper No. 7082 (1999).

——— "Why Public Schools Lose Teachers." *Journal of Human Resources* 39:2 (2004), 326-354.

Hedges, L.V., R. Laine, and R. Greenwald, "A meta-analysis of studies of the effects of differential school inputs on student outcomes." *Educational Researcher* 23:3 (1994), 5-14.

Lankford, H., S. Loeb, and J. Wyckoff, "Teacher sorting and the plight of urban schools: A descriptive analysis." *Educational Evaluation and Policy Analysis* 24:1 (2002), 37-62.

Latham, A. S., D. Gitomer, and R. Ziomek, "What the tests tell us about new teachers." *Educational Leadership* 56:8 (1999), 23-26.

Levinson, A., "Reexamining teacher preferences and compensating wages." *Economics of Education Review* 7:3 (1988), 357-364.

Nye, B., Konstantopoulos, S., and Hedges, L. "How Large Are Teacher Effects?" *Educational Evaluation and Policy Analysis* 26:3 (2004), 237-257.

Podgursky, M., "Defrocking the National Board: Will the imprimatur of 'Board Certification' professionalize teaching?" *Education Matter*, 1:2 (2001), 79-82.

Stone, J.E., "The value-added achievement gains of NBPTS-certified teachers in Tennessee: A brief report." College of Education, East Tennessee State University, 2:5 (2002).

Vandevoort, L. G., Amrein-Beardsley, A. and Berliner, D. C. "National board certified teachers and their students' achievement". Education Policy Analysis Archives, 12:46 (2004). Retrieved 11/23/05 from http://epaa.asu.edu/epaa/v12n46/.

Wilcox, D. D., "The National Board of Professional Teacher Standards: Can it live up to its promise?" In Marci Kanstoroom and Chester E. Finn, Jr. (Eds.), *Better Teachers, Better Schools*. (Washington, DC: The Thomas B. Fordham Foundation, 1999).

Wright, P., S. Horn, and W. Sanders, "Teachers and Classroom Heterogeneity: Their Effects on Educational Outcomes." *Journal of Personnel Evaluation in Education* 11:1 (1997), 57-67.

Table 1. Selected Sample Statistics[a]
(Standard deviations in parenthesis)

	Non-applicants		Applicant, Not NBPTS Certified		Applicant, NBPTS Certified	
	Reading	**Math**	**Reading**	**Math**	**Reading**	**Math**
Student Test Scores						
Post-test	149.47 (9.94)	150.39 (12.34)	149.47 (9.72)	149.80 (12.93)	151.52 (9.72)	152.38 (12.29)
Pre-test	143.78 (10.19)	140.64 (12.80)	143.65 (10.28)	140.67 (13.26)	145.34 (10.35)	142.17 (13.01)
Growth in test score in one year	5.69 (6.13)	9.75 (6.92)	5.83 (6.27)	9.14 (6.64)	6.18 (6.37)	10.21 (7.00)
Teacher Characteristics						
Black	0.14 (.34)	0.14 (.34)	0.24 (.43)	0.24 (.43)	0.07 (.25)	0.07 (.25)
White	0.85 (.35)	0.85 (.35)	0.75 (.41)	0.75 (.43)	0.93 (.26)	0.92 (.26)
Female	0.94 (.24)	0.94 (.24)	0.98 (.15)	0.98 (.15)	0.98 (.13)	0.98 (.13)
Years of teaching experience	13.18 (9.86)	13.18 (9.86)	12.54 (7.81)	12.52 (7.79)	12.55 (7.88)	12.56 (7.89)
Teacher Z-score	-0.03 (0.87)	-0.03 (0.87)	-0.16 (0.85)	-0.16 (0.85)	0.37 (0.87)	0.36 (0.86)
School Characteristics						
School of "excellence"[b]	0.02 (.13)	0.02 (.13)	0.02 (.14)	0.02 (.14)	0.07 (.25)	0.07 (.25)
School of "distinction"[c]	0.18 (.38)	0.18 (.38)	0.20 (.40)	0.20 (.40)	0.31 (.46)	0.31 (.46)
Proportion of free/ reduced-price lunch students	0.36 (0.19)	0.36 (0.19)	0.42 (.22)	0.42 (0.22)	0.31 (0.18)	0.31 (0.18)
Proportion of minority students	0.36 (0.25)	0.36 (0.25)	0.42 (0.38)	0.42 (0.28)	0.32 (0.22)	0.32 (0.22)
District/Community Characteristics						
Percent in community with at least a BA	16.92 (11.19)	16.92 (11.18)	15.25 (10.76)	15.23 (10.76)	19.52 (12.61)	19.51 (12.6)
Average household income (in thousands)	33.11 (6.33)	33.11 (6.34)	31.87 (6.59)	31.84 (6.59)	34.65 (6.53)	34.64 (6.53)
Median housing value (in thousands)	66.05 (16.52)	66.05 (16.52)	64.23 (16.59)	64.18 (16.60)	69.62 (17.75)	69.58 (17.75)
Sample size	600,261	602,577	4,602	4,622	4,297	4,318

[a] Student observations were divided into one of three categories: teacher was presently not applying to NBPTS ("non-applicant"); teacher was an unsuccessful current applicant that year ("Applicant, Not NBPTS Certified"); teacher was a successful current applicant that year ("Applicant, NBPTS Certified").
[b] Schools of excellence have 90–100% of their students performing at or above grade-level.
[c] Schools of distinction have 80–89% of their students performing at or above grade level.

Table 2. Selected Coefficients for Growth in Test-Score Models

(Standard errors in parenthesis)

	1	2	3	4	5	6	7	8	9	10	11	12
	Reading Achievement Growth						**Math Achievement Growth**					
Future applicant				.16*** (.04)						.21*** (.04)		
Current applicant			-.34*** (.09)	-.11 (.07)	-.37*** (.09)	-.57*** (.17)			-1.01*** (.10)	-.39*** (.07)	-1.06*** (.10)	-1.65*** (.19)
Past applicant				.08 (.09)						-.19* (.10)		
Future NBCT	.36*** (.05)	.35*** (.05)	.38*** (.05)		.34*** (.05)	.46*** (.09)	.62*** (.05)	.59*** (.05)	.69*** (.05)		.63*** (.06)	.66*** (.11)
Current NBCT	.18** (.08)	.16** (.08)					.08 (.09)	.04 (.09)				
New NBCT			.47*** (.13)		.48*** (.13)	.09 (.25)			1.10*** (.14)		1.17*** (.14)	1.59*** (.28)
Past NBCT			.22 (.14)		.27* (.15)	.04 (.27)			-.13 (.15)		-.11 (.16)	-1.22*** (.30)
Teaching exp: 1–2 yrs (base = 0 years)	.21*** (.04)	.21*** (.04)	.21*** (.04)	.21*** (.04)	.18*** (.04)	.34*** (.07)	.43*** (.04)	.43*** (.04)	.42*** (.04)	.42*** (.04)	.38*** (.04)	.41*** (.08)
3–5 yrs	.26*** (.04)	.27*** (.04)	.27*** (.04)	.27*** (.04)	.24*** (.05)	.44*** (.08)	.50*** (.05)	.51*** (.05)	.52*** (.05)	.50*** (.05)	.45*** (.05)	.54*** (.09)
6–12 yrs	.31*** (.04)	.32*** (.04)	.32*** (.04)	.31*** (.04)	.31*** (.05)	.54*** (.08)	.53*** (.05)	.55*** (.05)	.56*** (.05)	.54*** (.05)	.55*** (.05)	.76*** (.09)
13+ yrs	.42*** (.05)	.43*** (.05)	.43*** (.05)	.42*** (.05)	.43*** (.05)	.73*** (.09)	.63*** (.05)	.66*** (.05)	.66*** (.05)	.63*** (.05)	.66*** (.05)	.88*** (.10)
Teacher degree: master's (base=BA)	-.01 (.02)	-.02 (.02)	-.02 (.02)	-.01 (.02)	-.02 (.02)	-.01 (.04)	.07*** (.02)	.05*** (.02)	.06*** (.02)	.09*** (.02)	.05** (.02)	.10** (.04)
Doctorate degree	--.04 (.06)	-.05 (.06)	-.04 (.06)	-.03 (.06)	-.04 (.07)	-.25** (.12)	-.03 (.07)	-.05 (.07)	-.04 (.07)	<.01 (.07)	-.15** (.07)	-.40*** (.13)
Teacher license: full (base=other)	.22*** (.03)	.21*** (.03)	.22*** (.04)	.22*** (.03)	.20*** (.04)	.07 (.06)	.28*** (.04)	.28*** (.04)	.28*** (.04)	.30*** (.04)	.23*** (.04)	.25*** (.07)
Teacher Z- score	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Fixed-effect	No	No	No	No	School	Student	No	No	No	No	School	Student
R^2	0.06	0.06	0.06	0.06	0.07	0.57	0.10	0.10	0.10	0.10	0.11	0.57
Sample size	609,160						611,517					

***, **, *: Significant at 1%, 5%, and 10% confidence level, respectively.

Notes: These models also include controls for student's grade, race/ethnicity, gender, free/reduced-price lunch status (free/reduced-price lunch information was available in 1999 only), limited English proficiency status, learning disability status; teacher's age, race/ethnicity, gender, years of teaching experience, license type, license basis, degree level; total students at school, fraction of minority students at school, student-teacher ratio at school, fraction of free/reduced-price lunch students at school; total students in district, current per pupil expenditure, percent of education expenditure spent on instruction, urbanicity, starting salary for teachers with a B.A. and no experience in district; community's median housing value, percent in community with at least a B.A.; and the year of the test. Mean value replacement was used for cases where values for the explanatory variables were missing, except in the case of free/reduced-price lunch status, where missing values were coded as "no lunch information."

Table 3. Selected Coefficients for Growth in Test-Score Models Broken Out by Student Subgroup

(Standard errors in parentheses)

	Recipients of Free or Reduced-Price Lunch		Nonrecipients of Free or Reduced-Price Lunch		White		Black		Other	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
	Read	Math	Read	Math	Read	Math	Read	Math	Read	Math
Future NBCT	.55*** (.19)	1.09*** (.21)	.32** (.13)	.55*** (.14)	.34*** (.06)	.65*** (.06)	.31*** (.10)	.55*** (.12)	.83*** (.24)	.96*** (.27)
Current applicant	-.34* (.19)	-.86*** (.20)	-.20 (.17)	-.69*** (.18)	-.21* (.11)	-.71*** (.12)	-.41*** (.16)	-1.10*** (.18)	-.51 (.37)	-2.35*** (.41)
New NBCT	.98*** (.29)	1.52*** (.31)	.18 (.22)	.80*** (.23)	.36** (.15)	.89*** (.17)	.32 (.25)	.73*** (.28)	1.37** (.56)	3.10*** (.62)
Past NBCT	.58* (.30)	-.01 (.33)	.10 (.20)	.06 (.21)	.34** (.17)	-.02 (.18)	-.08 (.29)	-.24 (.33)	-.28 (.61)	-1.59** (.68)
Teaching exp: 1-2 yrs (base = 0 years)	.39*** (.10)	.93*** (.11)	.08 (.08)	.51*** (.09)	.17*** (.05)	.40*** (.05)	.22*** (.07)	.40*** (.08)	.53*** (.16)	.89*** (.18)
3–5 yrs	.62*** (.12)	1.03*** (.13)	.09 (.10)	.49*** (.11)	.17*** (.06)	.41*** (.06)	.32*** (.08)	.58*** (.09)	.67*** (.19)	.72*** (.21)
6–12 yrs	.59*** (.12)	1.05*** (.13)	.08 (.10)	.63*** (.11)	.22*** (.06)	.51*** (.06)	.46*** (.08)	.55*** (.09)	.53*** (.19)	.76*** (.21)
13+ yrs	.76*** (.13)	1.14*** (.14)	.17 (.10)	.56*** (.11)	.33*** (.06)	.60*** (.06)	.56*** (.08)	.77*** (.09)	.71*** (.19)	.68*** (.22)
Teacher degree: master's (base = BA)	.01 (.05)	.04 (.05)	.02 (.04)	.04 (.05)	.02 (.02)	.08*** (.02)	-.05 (.04)	-.01 (.04)	-.21*** (.08)	.07 (.09)
Doctorate degree	-.06 (.19)	-.03 (.21)	-.01 (.14)	-.02 (.15)	-.07 (.07)	-.08 (.08)	.09 (.13)	-.02 (.14)	-.36 (.29)	.35 (.33)
Teacher license: full (base = other)	.11 (.09)	.27*** (.10)	.26*** (.08)	.14* (.08)	.28*** (.04)	.32*** (.05)	.18*** (.06)	.36*** (.07)	.01 (.15)	-.10 (.16)
R^2	0.06	0.03	0.15	0.14	0.09	0.15	0.03	0.02	0.05	0.08
Sample size	85,177	85,970	111,927	112,133	400,369	401,434	175,466	176,502	33,194	33,447

***, **, *: Significant at 1%, 5%, and 10% confidence level, respectively.

Notes: The free/reduced-price lunch status models are run for records in 1999 only. These models also include controls for student's grade, race/ethnicity, gender, limited English proficiency status, learning disability status; teacher's age, race/ethnicity, gender, years of teaching experience, teacher Z-score, license type, license basis, degree level; total students at school, fraction of minority students at school, student-teacher ratio at school, fraction of free/reduced-price lunch students at school; total students in district, current per pupil expenditure, percent of education expenditure spent on instruction, urbanicity, starting salary for teachers with a B.A. and no experience in district; community's median housing value, and percent in community with at least a B.A. Mean value replacement was used for cases where values for the explanatory variables were missing, except in the case of free/reduced-price lunch status, where missing values were coded as "no lunch information."

The race models also include controls for student's grade, gender, free/reduced-price lunch status (free/reduced-price lunch information was available in 1999 only), limited English proficiency status, learning disability status; teacher's age, race/ethnicity, gender, years of teaching experience, teacher Z-score, license type, license basis, degree level; total students at school, fraction of minority students at school, student-teacher ratio at school, fraction of free/reduced-price lunch students at school; total students in district, current per pupil expenditure, percent of education expenditure spent on instruction, urbanicity, starting salary for teachers with B.A. and no experience in district; community's median housing value, percent in community with at least a B.A.; and the year of the test. Mean value replacement was used for cases where values for the explanatory variables were missing, except in the case of free/reduced-price lunch status, where missing values were coded as "no lunch information."

Table 4. Selected Coefficients for Growth in Test-Score Models Broken Out by Grade Level

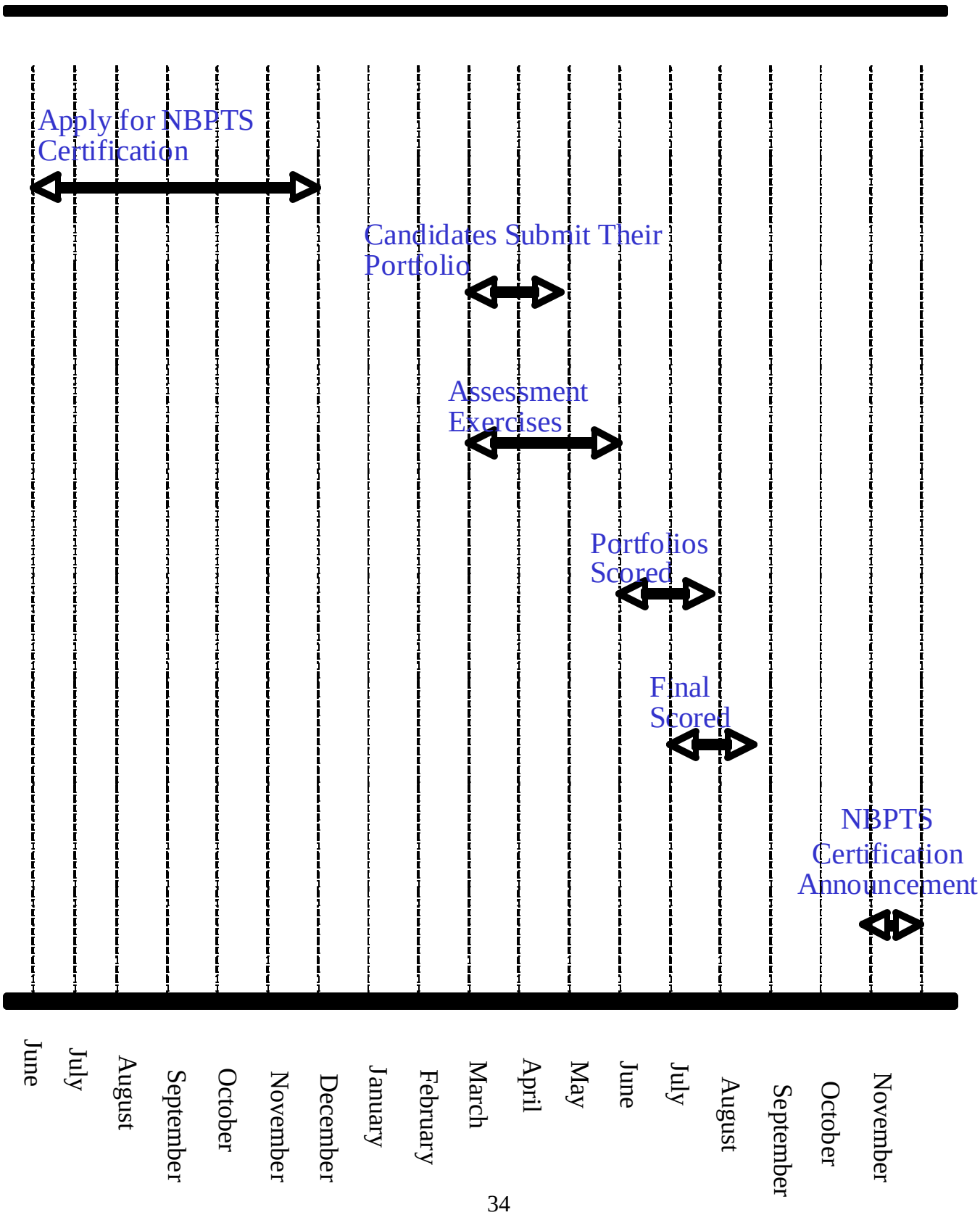
(Standard errors in parentheses)

	3rd Grade		4th Grade		5th Grade	
	1	**2**	**3**	**4**	**5**	**6**
	Read	Math	Read	Math	Read	Math
Future NBCT	.49*** (.09)	.95*** (.09)	.26*** (.08)	.34*** (.10)	.29*** (.08)	.51*** (.09)
Current applicant	-.38** (.16)	-1.07*** (.17)	-.26* (.16)	-1.18*** (.18)	-.44*** (.13)	-.64*** (.15)
New NBCT	1.01*** (.23)	1.72*** (.24)	-.10 (.22)	.94*** (.26)	.11 (.19)	.36* (.22)
Past NBCT	.63** (.31)	1.14*** (.32)	.27 (.23)	-.83*** (.27)	.03 (.20)	-.45** (.22)
Teaching exp: 1–2 yrs (base = 0 years)	.18** (.07)	.58*** (.08)	.29*** (.06)	.31*** (.07)	.17*** (.07)	.46*** (.07)
3–5 yrs	.22** (.09)	.60*** (.09)	.26*** (.07)	.51*** (.08)	.26*** (.07)	.48*** (.08)
6–12 yrs	.31*** (.09)	.72*** (.09)	.41*** (.07)	.54*** (.08)	.26*** (.07)	.43*** (.08)
13+ yrs	.38*** (.43)	.80*** (.09)	.53*** (.07)	.70*** (.09)	.35*** (.07)	.57*** (.08)
Teacher degree: master's (base = BA)	.03 (.03)	.09** (.04)	-.03 (.03)	.08** (.03)	-.05* (.03)	-.01 (.03)
Doctorate degree	.02 (.13)	-.23* (.14)	-.14 (.09)	-.30*** (.11)	.02 (.09)	.39*** (.10)
Teacher license: full (base = other)	.38*** (.06)	.50*** (.07)	.16*** (.06)	.28*** (.06)	.01 (.05)	.18*** (.06)
R^2	0.05	0.06	0.02	0.02	0.02	0.01
Sample size	228,654	229,623	191,853	192,606	188,653	189,288

***, **, *: Significant at 1%, 5% and 10% confidence level, respectively.

Notes: The grade-level models also include controls for student's gender, free/reduced-price lunch status (free/reduced-price lunch information was available in 1999 only), limited English proficiency status, learning disability status; teacher's age, race/ethnicity, gender, years of teaching experience, teacher Z-score, license type, license basis, degree level; total students at school, fraction of minority students at school, student-teacher ratio at school, fraction of free/reduced-price lunch students at school; total students in district, current per pupil expenditure, percent of education expenditure spent on instruction, urbanicity, starting salary for teachers with a B.A. and no experience in district; community's median housing value, percent in community with at least a B.A.; and the year of the test. Mean value replacement was used for cases where values for the explanatory variables were missing, except in the case of free/reduced-price lunch status, where missing values were coded as "no lunch information."

Figure 1. NBPTS Application Process Timeline



Appendix A

Variable Definitions [a]

STUDENT VARIABLES	DATA SOURCE	VARIABLE DEFINITIONS
Post-test	NCDPI	Student end-of-year test score
Pre-test	NCDPI	Student test score from end of previous year, unless student was a third grader. Third graders take a beginning-of-the-year test as pre-test.
Growth in test score in one year	NCDPI	Post-test score minus pre-test score
Grade (4), (5)	NCDPI	Student grade as recorded on state record
Race/Ethnicity (black), (other)	NCDPI	Student race as recorded on state record
Gender (female)	NCDPI	Student gender as recorded on state record
Year of record (1998), (1999)	NCDPI	Year of record
Free/reduced-price lunch status (free or reduced participant), (no lunch information)	NCDPI	Student's free and/or reduced-price lunch program eligibility (available from the state in 1999 only)
Limited English proficiency status	NCDPI	State indicates that student is classified as an LEP student
Learning disability status	NCDPI	State indicates that student has a learning disability in reading or math
TEACHER VARIABLES	**DATA SOURCE**	**VARIABLE DEFINITIONS**
Age	NCDPI	Teacher's age (in years) at start of each school year
Race/Ethnicity (black), (Hispanic), (Asian)	NCDPI	Self-reported race/ethnicity of each teacher
Gender (male)	NCDPI	Self-reported gender of each teacher
Years of teaching experience	NCDPI	Years of teaching experience that the state of North Carolina credits teachers with (can be for non-teaching but subject-related experience)
Teacher Z-score	NCDPI	Average Z-score from Z-scores of Praxis I and II, depending on what is available on teachers' record
License type (continuous license status)	NCDPI	Teacher is licensed on a permanent rather than temporary basis
License basis (North Carolina-based license)	NCDPI	Teacher received license through an education program approved by, and located in, the state of North Carolina
Master's degree	NCDPI	Teacher's highest degree is master's degree

Ph.D. /other advanced degree	NCDPI	Teachers highest degree is Ph.D. or other advanced degree
Future applicant	NBPTS[a]	Teacher has not yet applied to NBPTS, but we know applies in a future year for which we have data
Current applicant	NBPTS[a]	Teacher applies to NBPTS in current year
Past applicant	NBPTS[a]	Teacher applied to NBPTS in a prior year
Future NBCT	NBPTS[a]	Teacher becomes a NBCT in future year
Current NBCT	NBPTS[a]	Teacher is currently a NBCT
New NBCT	NBPTS[a]	Teacher becomes a NBCT in present year
Past NBCT	NBPTS[a]	Teacher had become an NBCT in a past year
SCHOOL VARIABLES	**DATA SOURCE**	**VARIABLE DEFINITION**
School of "excellence"	NCDPI web site	Schools of excellence have 90–100% of their students performing at or above grade-level, determined by state standards.
School of "distinction"	NCDPI web site	Schools of distinction have 80–89% of their students performing at or above grade-level, determined by state standards.
Total students	Common Core of Data	Total number of students enrolled at school
Fraction of minority students	Common Core of Data	The fraction of minority students in school out of the total school population
Student-teacher ratio	Common Core of Data	Total students in school divided by total number of full-time equivalent teachers
Fraction of free/reduced-price lunch students	Common Core of Data	The fraction of students eligible for free/reduced-price lunch programs under the National School Lunch Act of the total school population
DISTRICT/COMMUNITY VARIABLES	**DATA SOURCE**	**VARIABLE DEFINITION**
Total students	Common Core of Data	Total number of students enrolled in PK–12 grades in district's schools
Current per pupil expenditures ($)	Common Core of Data	Current expenditures are expenditures for the day-to-day operation of schools and school districts. They include expenditures for instruction, support services, food services, and enterprise operations. They exclude expenditures for capital outlays and programs outside the regular preschool to grade 12 scope.
Percent education expenditure spent on instruction	Common Core of Data	Instruction expenditures divided by total expenditures
Urbanicity (urban, suburban)	Common Core of Data	This is a composite of local codes from the schools. Urban districts are those in large or mid-size central cities. Suburban

		districts are found in the urban fringe of large or mid-size central cities. Rural districts are defined as those in large towns, small towns, or census-defined rural areas.
Starting salary with BA and no experience	NCDPI web site and phone calls to individual school district offices	Starting salary for teachers with a bachelor's degree and no experience and in that district
Percent in community with at least a BA	Common Core of Data	Percent of residents with a bachelor's degree or higher degree in community
Average household income (in thousands)	Common Core of Data	Average income for a household of four in the community
Median housing value (in thousands)	1990 Census data	Median value of all housing units in district. Value is the respondent's estimate of how much the property (house and lot, mobile home and lot, or condominium unit) would sell for if it were for sale.

[a] For NBPTS variables, data are only available for teachers who applied for National Board Certification.

Appendix B

Table B.1. Student and Teacher Record Matches with Pre- and Post-Tests

	1997		1998		1999		Total	
Total student records	286,574		296,609		306,469		889,655	
Student and teacher merged records (STMs)	246,049		256,840		268,648		771,537	
STMs w/ end-of-grade (post) test scores	*Read*	*Math*	*Read*	*Math*	*Read*	*Math*	*Read*	*Math*
	236,462	236,914	245,542	246,607	257,085	258,323	739,089	741,844
STMs w/valid pre- and post-test scores	*Read*	*Math*	*Read*	*Math*	*Read*	*Math*	*Read*	*Math*
	202,443	202,886	190,839	191,605	215,878	217,026	609,160	611,517

Table B.2. Number of Teachers in Each NBPTS Category by Year

READING - Teacher Observations								
	Future applicant	**Current applicant**	**Past applicant**	**Future NBCT**	**Current NBCT**	**New NBCT**	**Past NBCT**	**Total**
1997	536	27	17	335	11	11	0	937
1998	469	146	27	280	77	65	12	1,076
1999	340	285	169	190	215	140	75	1,414
Total	1,345	458	213	805	303	216	87	3,427

MATH - Teacher Observations								
	Future applicant	**Current applicant**	**Past applicant**	**Future NBCT**	**Current NBCT**	**New NBCT**	**Past NBCT**	**Total**
1997	538	27	17	335	11	11	0	939
1998	471	146	27	281	77	65	12	1,079
1999	340	285	169	190	215	140	75	1,414
Total	1,349	458	213	806	303	216	87	3,432

Table B.3. Numbers of NBCTs/Students with NBCTs in North Carolina across Years and Grades

READING	1997	1998	1999	Total
Grade 3	4 / 80	21 / 415	84 / 1,722	109 / 2,217
Grade 4	3 / 62	24 / 481	65 / 1,226	92 / 1,769
Grade 5	4 / 89	32 / 609	66 / 1,345	102 / 2,043
Total grade 3–5 NBCTs	11 / 231	77 / 1,505	215 / 4,293	303 / 6,029

MATH	1997	1998	1999	Total
Grade 3	4 / 81	21 / 420	84 / 1,729	109 / 2,230
Grade 4	3 / 62	24 / 482	65 / 1,225	92 / 1,769
Grade 5	4 / 90	32 / 610	66 / 1,349	102 / 2,049
Total grade 3–5 NBCTs	11 / 233	77 / 1,512	215 / 4,303	303 / 6,048

Endnotes

[1] Specifically, NBPTS was founded in 1987 with a threefold mission: (1) to establish high and rigorous standards for what accomplished teachers should know and be able to do; (2) to develop and operate a national voluntary system to assess and certify teachers who meet these standards; and (3) to advance related education reforms to capitalize on the expertise of National Board Certified Teachers (NBCTs) (http://www.nbpts.org).

[2] http://www.nbpts.org/pdf/quickfacts.pdf, accessed 11/22/05.

[3] Through September 2005. NBPTS states that a portion of these federal funds have provided subsidies to candidates to fund a portion of their respective candidate fees. (http://www.nbpts.org/pdf/quickfacts.pdf, accessed 11/22/05).

[4] Due to variations across years in the available data, this estimate was calculated in several steps. The application fee was originally $1,800 (through 1998), and increased to $2,300 beginning in 1999. Since the actual number of NBPTS applicants prior to 1999 is not available due to problems with the data, we estimate the number of applicants from 1993-98 by multiplying the number of NBCTs (http://www.nbpts.org/nbct/nbctdir_byyear.cfm) in those years by 2 (according to ETS officials, the average first-time pass rate is roughly 50 percent). This gives us an estimated 9,612 applicants for the pre-99 years, which we multiply by $1,800, for a total estimated cost of $17.3 million for the first six years of the program. We then multiplied the actual number of applicants for 1999–2004 (91,316, according to an e-mail to the author from Anika Sandy-Hanson of NBPTS on 11/22/05) by $2,300 for a total actual cost of $210 million for 1999–2004. Adding these two amounts together gives us our total estimated application fees of $227.3 million.

[5] As of September 2005, legislative and policy action creating incentives and recognition for National Board Certification has been enacted in all 50 states and in approximately 544 local school districts, including the District of Columbia; many of these offer at least one type of financial incentive (bonus or salary supplement) for teachers to become NBPTS-certified (http://www.nbpts.org/about/state.cfm, accessed 11/2/05).

[6] See http://www.cde.ca.gov/ta/sr/nb/faqs.asp for more information on this NBPTS incentive.

[7] For information on the NBPTS process as professional development, see http://www.nbpts.org/standards/nbcert.cfm. For a detailed description of the assessment see http://www.nbpts.org/standards/dev.cfm.

[8] See, for instance, Hanushek (1986, 1997) who suggests there is little relationship between teacher characteristics and student outcomes, and Hedges et al. (1994) or Card and Krueger (1996) who find more positive results. We discuss "teacher effectiveness" in this paper in terms of the teachers' contributions towards student gains in achievement.

[9] Education productivity studies typically measure the size of the relationship between various quantifiable education factors and student achievement. Goldhaber et al. (1999), for example, investigate the contributions of school, teacher, and class characteristics on student achievement. They find only about 3 percent of the contribution teachers make toward explaining student achievement is associated with teacher experience, degree level, and other readily observable characteristics. The remaining 97 percent is made up of teacher qualities or behaviors that could not be separately isolated and identified.

[10] The first-time pass rate was 48 percent during that time, according to ETS and NBPTS officials. It is important to distinguish between the first-time pass rate and individuals' overall pass rate: individuals' overall pass rate will be higher than the first-time pass rate, since individuals who are not initially successful in obtaining certification may re-apply to NBPTS in later years.

[11] This is the percentage of candidates who passed the Praxis II exam from 1994 to 1997, which is the most commonly used standardized test for teacher licensure (Latham et al. 1999).

[12] We recognize that test performance is just one of many ways that NBCTs might influence student outcomes, and that NBPTS certification may have numerous impacts on students. For instance, NBCTs might affect the propensity of students to drop out of school or pursue interests in particular subjects.

[13] In addition to this specification we estimated lagged test-score models where the post-test score is regressed on a lagged test-score along with the other controls specified in equation 1, and we estimated models that included school fixed-effects, teacher fixed-effects, or student fixed-effects.

[14] We have information on this variable for 1999 only, when the state started collecting it.

[15] For information on any variables in our data set, see appendix A.

[16] http://www.nbpts.org/about/govt_nochild.cfm, accessed 11/2/05.

[17] Unfortunately, as we discuss in greater detail below, we are somewhat limited in our ability to estimate the effects of teachers who either became NBPTS-certified or applied to the program but failed to gain certification in the years following application; consequently there is also little variation to allow for the estimation of the NBPTS variables in a teacher fixed-effects specification. The reason is that the number of NBPTS applicants expanded rapidly during the years of our data, so the number of applicants and certified teachers is significantly larger in the final year of the data than in any other year, implying few past applicants and certified teachers.

[18] There are 74,318 teachers in 1997, 76,609 teachers in 1998, and 78,075 teachers in 1999.

[19] For most of the years of data in our study, more than one-quarter of the nation's NBCTs taught in North Carolina.

[20] State officials (a statistical analyst at the NCDPI testing office and that individual's supervisors) stated that at least 90 percent of the time, the students' classroom teacher is the same person as the one listed on their student record as their "test administrator." We followed up on that information, asking the same question of district-level testing officials. We gathered our sample by calling the first 15 districts (in alphabetical order) as well as the 5 districts serving the largest metropolitan areas in the state, until we received 10 districts' responses on the question. All 10 responses confirmed that at least 90 percent of the time, the testing administrators were indeed also students' classroom teachers (a list of the districts that provided this information is available on request). To minimize any mismatching of student and teacher records, we excluded magnet schools from our sample, since state officials indicated that there is a reduced likelihood at magnet schools that the teacher listed as the testing administrator on the student's record is in fact that student's regular teacher.

[21] We retained only teacher test scores that fall into the proper range for each particular test (e.g., if the range of possible scores for a test was between 100 and 200, and the recorded score was a 54, we considered the test to be missing). We replaced missing teacher test scores with the mean value for that particular test in order to keep as many observations in our models as possible.

[22] For teachers with multiple scores for the same test, the most current test score was used.

[23] The most commonly available tests on teacher records are the Praxis II tests, because the state of North Carolina began requiring teachers to take and submit their Praxis II exam scores in the second half of the 1990s in order to obtain a teacher's license. Teachers have until June 30 of the year that their teaching license is granted to take and report their Praxis II exams. Mean value replacement was used when teachers were missing test scores.

[24] Please see **figure 1** in appendix A for a graphical representation of the NBPTS application and certification timeline. It should be noted that the length of the entire process varies for candidates, depending on when they apply (June through December in the year previous to the fall that certification is announced).

[25] Whether a teacher is considered certified upon completion of the assessment requirements or after certification announcements have been made could impact the results of the data analyses, since it is thought the assessment process can in fact detract from a teacher's performance due to its rigor and time commitment.

[26] For more information on how various growth/gain goals and measures are determined by the state, see http://www.ncpublicschools.org/accountability/reporting/abc/2002-03/?&print=true (11/2/05).

[27] For specific information on how student and teacher records were merged, please contact the authors.

[28] Some students and teachers may appear in our data in multiple years, so we refer to each appearance as an "observation." Some student and teacher do not appear in our data, despite the presence of their records. We did not include observations where we were unable to match teachers to students. This happened primarily because of coding errors in listing the teacher

name on the student's record. For instance, since we matched student and teacher records by teacher name and school code, if teachers' last names or first names were spelled differently on each year's records, it was not possible to match the information. Matching students across years is similarly problematic if data is entered differently between years, making it impossible to match multiple-year records of the same student in those cases.

[29] Of the reading observations, 6,029 student observations have a new or past NBCT and 303 teacher observations are new or past NBCTs in our data. Of the math observations, 6,048 student observations have a new or past NBCT and 303 teacher observations are new or past NBCTs in our data. See **table B.3** for the distribution of these students and teachers across years and grades.

[30] Unless otherwise noted, the effect sizes reported in this paper are based on the standard deviations of the growth in reading (6.14) and math (6.91) scores, for the full sample of students.

[31] Though we do not report them all here, a complete set of coefficient estimates is available from the authors upon request.

[32] All else equal, in reading and math, the one-year test-score growth for white students exceeds black students by 10 and 12 percent of a standard deviation respectively; male students exceed female students by 2 and 1 percent of a standard deviation, respectively; nonparticipants in the free and reduced-price lunch program exceed participants by 4 and 9 percent of a standard deviation, respectively; and students without learning disabilities exceed students with learning disabilities by 5 and 8 percent of a standard deviation, respectively.

[33] See Hanushek (1986, 1997) and Greenwald et al. (1996) for reviews of this literature.

[34] The elimination of special education and limited English dependent students from the sample reduces the sample size to 565,210 for reading and 565,913 for math. The only changes in results for NBPTS variables, reported in **table 2**, occur in the reading models: In the reduced sample, current NBCT is only significant at the 10 percent level in the reading models displayed in columns 1 and 2 of table 2, and current applicant is significant at the 10 percent level in column 4 of table 2.

[35] We are able to match teachers and students through the 1998–99 school year, but we have data on teacher certification status through 1999–00. Since we have one more year of information on teacher NBPTS certification status than student achievement, our sample of future NBCTs (805 unique teacher observations) is significantly larger than our sample of current NBCTs (303 unique teacher observations).

[36] One might argue that it is more appropriate to use either the standard deviation of student *gains* or the standard deviation of *teacher effects* (both of which are smaller) as a benchmark to measure the size of the NBPTS effects.

[37] The decrease in the magnitudes of current NBCT and future NBCT is not statistically significant in either the reading or math models.

[38] A *t*-test of the difference in the future NBCT and past NBCT coefficients indicates that this difference is significant in the math model.

[39] In our sample, there are only 87 unique teachers who are past NBCTs.

[40] The "Ashenfelter dip" refers to the possibility that estimates of job training programs are biased by the fact that trainees typically experience a decline in their earnings (the dip), relative to comparison groups, just prior to enrollment in a training program.

[41] The teacher fixed-effects models account for the possibility that unobservable variables at the teacher level might be correlated with teachers' NBPTS status.

[42] See appendix **table B.2** for details on the number of teachers who fall into each NBPTS category.

[43] A similar pattern exists when we focus on those who are actually certified rather than all applicants.

[44] We estimated a specification where the post-test score (as opposed to the gain) was regressed against the set of controls and school fixed effects and find that only past NBCT changes: it is significant at the 1 percent level in this model, as opposed to being significant at the 10 percent level in the gains model.

[45] In addition to this, we estimated a student fixed-effects specification where the post-test score (as opposed to the gain) score was regressed against the set of controls. The coefficient estimates on the NBPTS variables were quite similar to those reported in columns 6 and 12 of **table 2**. The only difference in the estimates occurs on the estimated coefficient of new NBCT (see column 6 of **table 2**), which is positive and statistically significant at the 5 percent level in the reading model, whereas it was not statistically significant in the gain specification (though it was positive).

[46] We also estimated more restrictive specifications of the model that included interactions between NBPTS status variables and race/ethnicity and lunch status. The results from these models reflect the findings we discuss below.

[47] The "other" race variable consists of students reported to be Asian, Hispanic, Native American, mixed, and other in the state data.

[48] Furthermore, the sum of current applicant and new NBCT is positive and statistically significant in the 3rd grade reading and math models, implying that having a NBCT *in the year of application* is beneficial for students in that grade.

[49] One must interpret this finding with caution, as it does not necessarily imply that schools would be better off *reassigning* their NBCTs to teach lower-income students or in earlier grades, since teaching skills may not easily be transferable across grades or types of students.