THIS IS AN UPDATED VERSION OF A PAPER OF THE SAME TITLE PREVIOUSLY POSTED ON THE URBAN INSTITUTE WEBSITE ON 3/8/04.

Can Teacher Quality Be Effectively Assessed?

Dan Goldhaber
(University of Washington and the Urban Institute)

Emily Anthony
(Urban Institute)

April 27, 2004

Abstract

In this paper, we describe the results of the first large-scale study, based on a unique data set from North Carolina, assessing the relationship between the certification of teachers by the National Board for Professional Teaching Standards (NBPTS) and elementary- level student achievement. Our findings indicate that NBPTS is successfully identifying the more effective teachers among applicants, and that NBPTS-certified teachers, prior to becoming certified, were more effective than their non-certified counterparts at increasing student achievement. The statistical significance and magnitude of the "NBPTS effect," however, differs significantly by grade level and student type.

Acknowledgements —

We 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 seminar participants at the Peabody College at Vanderbilt University's Department of Leadership, Policy, and Organizations. 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, as well as 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 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 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 outside of 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]

Participation in the NBPTS program has grown dramatically over a relatively short period of time. The National Board began by certifying fewer than 100 teachers in 1994–95, and by November 2003 had certified approximately 32,000 teachers (out of approximately 65,000 applicants[2]). This dramatic increase in current applicants and National Board Certified Teachers (NBCTs) is likely to be at least partially attributable to the incentives that many states and districts have adopted for NBPTS certification. Many pay at least a portion of the $2,300 application fee required for the NBPTS assessment. The total value of fees paid to NBPTS (by localities, states, or teachers) is approximately $150 million.[3] This is on top of the considerable direct federal (over $100 million) and private (over $100 million) support that NBPTS has received.[4] In addition, some states and localities offer salary supplements to NBCTs.[5] In North Carolina, for instance, NBCTs receive a 12 percent increase in their base pay, and in California, NBCTs who opt to teach in a high-poverty school for four years were at one time 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.

The recognition and financial incentives that NBPTS certified teachers receive in various states and school districts indicate that many policymakers view this certification as a signal of teacher quality. The NBPTS assessment process may also be viewed as an important professional development opportunity for teachers (http://www.nbpts.org/standards/nbcert.cfm), since it involves a number of exercises designed to require "intense self-reflection and analysis" of their own teaching,[7] which, at least in theory, has the potential to build human capital. National Board advocates view NBPTS as an important vehicle for setting high standards (and raising them) for classroom educators, for "professionalizing" teaching, and for encouraging positive education reforms overall—ultimately contributing to the goal of improving student achievement.

National Board skeptics, by contrast, view NBPTS as an organization that has garnered significant public and private investment despite a lack of evidence on its efficacy in identifying effective teachers (Finn 2003). In fact, a survey of the literature reveals surprisingly little quantitative evidence on whether NBPTS is successfully accomplishing its stated mission: to advance the quality of teaching and learning by establishing standards for what accomplished teachers should know and be able to do, and recognizing those teachers who demonstrate mastery of those standards. It remains an open question as to whether this certification should be treated as a signal of teacher quality, or if the process should be viewed as one that builds the human capital of teachers.

In this paper, we describe the results of a study assessing the relationship between NBPTS certification of teachers and elementary-level student achievement. More specifically, using a unique data set from North Carolina, we estimate student level value-added models and test whether the value added by NBCTs differs from that of unsuccessful current applicants and non-applicant teachers. Our findings indicate that NBPTS is successfully identifying the more effective teachers among applicants, and that NBPTS-certified teachers, prior to becoming certified, were more effective than their non-certified counterparts at increasing student achievement. The statistical significance and magnitude of the "NBPTS effect," however, differs significantly by grade level and student type.

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 some conclusions and policy recommendations.

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, Brewer, and Anderson 1999; Hanushek, Kain, and Rivkin 1999; Wright, Horn, and Sanders 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 will achieve 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] For example, the teacher attributes commonly used for certification, recruitment, screening, and selection of teachers (i.e., certification status, degree, and experience levels) are not strongly correlated with student learning gains (Goldhaber and Brewer 2000; Hanushek 1986, 1997). In other words, teachers clearly matter, but *teacher quality* is not strongly related to observed *teacher credentials.* 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] Consequently, it may be necessary to assess what teachers are actually doing in the classroom, not simply check their credentials, in order to evaluate teacher quality.

NBPTS was founded upon the notion that the attributes that make experienced teachers successful can, in fact, be measured based on applicants' ability to demonstrate mastery of a set of standards laid out by the National Board.[10] Based on the overall NBPTS pass rate (described below), NBPTS certification appears to be more difficult to attain than the standard hurdles associated with licensure testing used in most states. For example, from 1999 through 2002, approximately 50 percent of first-time applicants to the National Board became certified[11], a

much lower success rate than the national average of nearly 90 percent of teachers who pass a licensure exam on their first try.[12]

Research (Goldhaber, Perry, and Anthony 2004) on those who apply to and are certified by NBPTS finds a strong correlation between teacher performance on standardized exams (e.g., licensure tests) and both the probabilities of application and, given application, NBPTS certification. This relationship is important, since numerous studies find a positive connection between teacher performance on measures of academic proficiency and student outcomes (Ferguson and Ladd 1996; Goldhaber 2002; Greenwald, Hedges, and Laine 1996). As previously discussed, however, there is a dearth of evidence relating NBPTS certification to a direct measure of teacher effectiveness: student outcomes. In fact, National Board skeptics often note that NBPTS lacks sufficient research to back its standards, charging that the process is based on internal validity, frequently measured against the National Board's own standards of appropriate teacher practices, rather than external validity measures of student achievement (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. We are aware of only two studies (Bond et al. 2000 and Stone 2002) that attempt to link NBPTS certification status directly to student outcomes. Both are based on relatively small samples of teachers and students, and they reach divergent conclusions regarding the ability of NBPTS to identify the more effective teachers.[13] Until now, 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 voluntary NBPTS teacher certification program, despite its powerful potential to identify

teaching skills that may relate to student learning.

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.[14] We begin by estimating a basic educational production function of the following form:

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

The subscripts i, j, s, and t denote individual student, teacher, school/district/community, and time, respectively. The left hand side of the equation $A_{ijst} - A_{ijs(t-1)}$ 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 a student's race, gender, learning disability, free or reduced-price lunch status,[15] 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, 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 students receiving 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.[16]

The NBPTS vector contains our main variables of interest. An ideal specification would include variables identifying future, current, and past application to NBPTS and future, current,

and past certification by NBPTS. Unfortunately, the number of certified teachers is significantly larger in the final year of our data than in any other year, implying that the sample contains few past applicants and certified teachers. Thus, we experiment instead with a number of specifications that include information on future and current applicants and NBCTs, which we can use to answer several questions about the relative impact of teachers by NBPTS certification status.

A comparison of NBCTs to non-certified teachers is essential for policymakers wishing to use the NBPTS credential as a signal of teacher quality. This credential is actually cited in the federal No Child Left Behind Act as a prime example of the ways in which teachers can meet its "highly qualified" requirement, and which many states are incorporating into their regulations as meeting this federal requirement (http://www.nbpts.org/about/govt_nochild.cfm). Given the $2,300 per teacher NBPTS assessment cost and previous research showing a correlation between teacher test performance and the probability of certification, it is important to ask whether one might be able to identify highly effective teachers simply from their test (e.g., licensure) performance. A unique aspect of the North Carolina data used in this study is that it gives us a measure of teacher academic proficiency (mainly teachers' performance on licensure exams) for the great majority of teachers in our sample. Thus, we are able to assess the degree to which NBPTS certification conveys information about a teacher's effectiveness above and beyond that conveyed by a teacher's performance on licensure tests.

While interesting, the comparison of certified and uncertified teachers does not provide evidence on whether the NBPTS assessment system is accurately identifying highly effective teachers. NBPTS is making judgments only about those teachers who have applied for certification, so a comparison between current NBCTs and non-certified teachers ignores the

possibility that the NBPTS applicant pool might be very different from the teacher workforce as a whole. For example, in the hypothetical case in which all applicants are more effective teachers than non-applicants, we would always observe a positive effect of NBCTs regardless of how accurately the certification process identified effective teachers. The converse is also true. Thus, to judge the effectiveness of the NBPTS assessment process, we compare NBCTs to unsuccessful NBPTS applicants.

The NBPTS assessment process is also viewed by many as a means of adding to the human capital of teachers. Thus, we might expect NBCTs to be more effective teachers after they have gone through the assessment. We can test this by including a variable identifying a teacher's future certification status to show whether those teachers who are eventually NBPTS certified were more effective teachers prior to receiving their certification. The differential between this coefficient and the coefficient identifying current NBCTs helps to answer the question of whether going through the NBPTS certification process actually adds to teachers' human capital or whether NBPTS is simply a screening device that signals teacher quality.

Finally, we might be concerned that NBCTs have differential impacts on different types of students, or that our findings are confounded by non-random sorting of students across teachers (Clotfelter, Ladd, and Vigdor 2003). In order to disentangle the relationship between student achievement growth and teacher attributes, we 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.[17] 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.[18] Furthermore, it is possible with these data to link teacher and student records (at the elementary level) and to track both over time.[19]

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.[20] 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.[21] We used 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.[22]

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.[23] Based on the NBPTS application and certification timeline (shown in **figure 1** in 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.[24]

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 1998–1999 only), English proficiency status, grade and year, and test results for grades three through ten. 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.[25]

We opted to restrict our study to elementary students in the third, fourth, and fifth 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.[26] 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 third grade, or the end of the previous year for the fourth and fifth grades).[27] 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 over the 1996–97 school year to the 1998–1999 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.[28]

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 had a teacher going through the NBPTS assessment process that year. Approximately 6,000 students in our sample have teachers who were successfully certified by NBPTS by the time that teacher taught them. Although students with teachers certified, on average, have 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 one-percent level) than the growth for both those who have non-applicant 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 one-percent level). These differences, nonetheless, are relatively small; the largest differential is in math between certified and non-certified teacher applicants, at just over a point on the exam or roughly 14 percent of a standard deviation in the growth in math scores.[29]

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 judged as high performing, with fewer disadvantaged students. Furthermore, NBCTs themselves, both applicants and non-applicants, 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

Table 2 shows coefficient estimates for key NBPTS variables for our reading and math achievement models. Columns 1 through 6 show the results for various specifications of the growth in the reading score model, and columns 7 through 12 show various specifications for the growth in the math score model. Prior to discussing the coefficient estimates for the three NBPTS dummy variables, it is worth noting that many of the individual student and schooling variables are statistically significant.[30] We find that 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, non-participants in the free/reduced-price lunch program, and/or who do not have learning disabilities.[31]

The signs and statistical significance levels for many of the schooling variables are mixed and sensitive to model specification, which is consistent with much of the educational productivity literature.[32] For example, years of teaching experience and having a master's degree and a "continuous" teaching license from the state (as opposed to a provisional or temporary license or one that only meets initial teaching license requirements for the state) are generally positive and significant. In some cases, however, we also find that larger classes and schools

seem to benefit students, which seems counterintuitive, although it is consistent with many of the findings on class size in the educational productivity literature (Hanushek 1986).

A. Measures of NBCT Effectiveness

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 their 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 1999–00 school year ("Future NBCT").[33] The excluded comparison group in this specification (non-NBCTs) includes those teachers who are either non-applicants or who apply to the program but are unsuccessful in achieving certification.

The magnitudes of the Future NBCT coefficients suggest that student gains produced by the teachers who are certified by NBPTS exceed those of non-certified 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). 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 1996).

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 t**able 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.[34] 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 proportion 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 the table 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 non-certified 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 attempts 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 suggest that students with teachers who are certified would be expected to achieve growth that exceeds that of students with teachers who were 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 attempts at certification are actually less effective than non-applicant teachers (this effect is identified by the coefficient on Current Applicant).[35] 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 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. Our reading results do not support this hypothesis, however, as the magnitude of the Past NBCT coefficient is smaller (and not significant) than the Future NBCT coefficient, and in the math model, is actually

negative (but not significant). We interpret these last findings with caution since there are very few teachers who fall into the Past NBCT category in our data set.[36]

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

If it were it the case that going through the NBPTS assessment process makes a teacher more effective, we might expect to see a significantly larger coefficient on the Past NBCT variable than on New NBCT or Future NBCT, indicating that teachers are more effective in the years after they completed the assessment than they are prior to attempting certification. As we described above (in reference to columns 3 and 9 of **table 2**), the magnitudes of these coefficients do not support this idea, since NBCTs appear to be more effective before they are certified than after. However, as we note above, given the relatively small number of Past NBCTs in our dataset, we want to be careful about drawing strong conclusions based on this model specification.

We further explore this issue by estimating specifications of the model that test whether *any* applicant—either successful or unsuccessful—benefits from going through the assessment process. Specifically, we estimate a model specification (reported in column 4 of **table 2** for reading, and column 10 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 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 prior or post-application year. The test of human capital effects, however, provides mixed evidence on the impacts of the assessment. The results for the reading models are consistent with our NBCT findings reported above in that NBPTS applicants as a whole do not appear to be more effective after they have gone through the assessment process than they were pre-assessment (the coefficients on Past Applicant are not greater in magnitude than the coefficients on Future Applicant in the reading model). The math models' results, by contrast, do lend support to the notion that the assessment process adds to teacher human capital: the coefficient on Past Applicant is significantly larger than on Future Applicant, implying that teacher applicants are more effective in math after they have completed the NBPTS assessment process.[37]

However, we wish to be cautious about these results as well, since they may simply reflect effects associated with particular cohorts attempting certification.[38] 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.[39]

C. School and Student Fixed Effects Model Specifications

There is an important reason to interpret all of the above findings with caution. 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, Loeb, and Wyckoff 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, Kain, and Rivkin (forthcoming) find 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 and Cramer (2003) and Clotfelter, Ladd, and Vigdor (2003) suggest that these non-random sorting patterns may be even more pronounced for NBCTs.

If teachers and teacher quality are non-randomly 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 teacher characteristics are likely biased. In particular, the impact of teacher credentials would be overstated if the most motivated, high-achieving students tend to be assigned to more highly credentialed teachers. In fact, recent empirical research on teachers and students from North Carolina shows that the impact of teacher credentials does tend to be overstated in simple

statistical models precisely because of the distribution of teacher credentials across student characteristics (Clotfelter, Ladd, and Vigdor 2003).

To take account of the non-random distribution of teacher credentials 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 the table shows the results for math growth. The inclusion of these school effects makes surprisingly 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 non-certified 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 non-random distribution of NBCTs across schools.[40]

The above models account for non-random sorting of teachers across schools, but they do not account for the possibility that teachers may be non-randomly 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. 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 teachers. Specifically, in reading, newly certified teachers do not appear to be more effective than non-certified 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 teachers are actually less effective than are non-applicants. Again, these results must be interpreted with some caution, given that we are estimating these models 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.

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 all the way back to the "Coleman Report" (Coleman et al. 1966) tends to find that teacher quality has a larger impact on lower achieving students than those with higher achievement. Furthermore, NBPTS is thought to value particular constructivist approaches to teaching (Wilcox 1999; Ballou 2003) that may be more or less effective when applied with different types of students or to students of varying ages or academic achievement level. 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**).

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 (.85 in reading and .97 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 .56 in math) (see **table 3**).[41] 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 NBCT, some positive NBPTS identification effects but little or no difference between NBCTs and non-applicants in the year of application, and divergent results for math and reading for Past NBCT.

We also observe important differences in the estimated effects by grade level. In general, the largest significant positive NBPTS effects are found in the third 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 third grade than in either the fourth or fifth grades (comparing across columns in **table 4**). The New NBCT coefficient is significant and positive for both reading and math in the third grade, but outside of the third grade it is only positive and significant for fourth grade math.[42]

In contrast to the overall findings, the third grade results and the fourth grade math model results provide at least some evidence that successfully completing the certification process itself does, in fact, make teachers more effective, as the magnitude of the coefficient on New NBCT is larger than the magnitude of the coefficient on Future NBCT. Again, this result must be treated with caution as it may simply reflect cohort differences.

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. We discuss this and other public policy issues in greater detail in the following section.

V. Public Policy Implications: The Costs of Identifying High-Quality Teachers

So what are the policy implications of our findings? To begin with, this is the first large-scale study that appears to confirm the NBPTS assessment process is effectively identifying those teachers who contribute to relatively larger student learning gains. This finding is important both because it provides some indication of a positive return on the investment in NBPTS, and on a more fundamental level, it demonstrates that it is actually possible to identify

teacher effectiveness through NBPTS-type assessments. The NBPTS process is not, however, an inexpensive one. The direct cost to the state of North Carolina of identifying one NBCT (in 2000 dollars) totals more than $6,500: $2,300 for the assessment itself, along with a salary increase of about $4,200, which is 12 percent (the state salary supplement) of the average NBCT salary in 2000 (Goldhaber and Anthony 2003). The total cost of identifying an NBCT is more than $8,800, based on a pass rate of about 50 percent.[43] This is actually a lower-bound estimate, given that teachers receive a salary supplement in each year for the 10-year life of the certificate.

Is the investment in the NBPTS process a cost-effective one? The answer depends on what type and grade level of students are being taught by NBCTs as well as by those who are unsuccessful in their attempt to become certified.[44] It is also of significant policy import that unsuccessful Current Applicants are actually less effective teachers in the year in which they apply to NBPTS. These findings suggest that the total impact of the NBPTS program on schools depends both on the number of successful and unsuccessful applicants in a given school, as well as where those teachers are assigned. Consequently, schools with many unsuccessful applicants or those with successful applicants that leave following their certification may actually be worse off for having had their teachers apply to the program, since NBCTs are no more (or are less) effective than non-applicants in the year of application and unsuccessful applicants are less effective in the year of application. Teaching assignment matters because schools with NBCTs receive substantially more educational benefits from having their NBCTs teach low-income students in earlier grades. However, this does not necessarily imply that schools would be better off reassigning their NBCTs to teach lower-income students or third graders, as teaching skills may not easily be transferable across grades or types of students.

Using the most optimistic NBPTS findings on human capital (i.e., that the assessment process adds about .4 test score points in math, which is roughly 3 percent of a standard deviation), we can loosely compare the benefits of investments in the NBPTS model to estimates of the benefits of other educational interventions.[45] The cost of the assessment is $2,300 for all applicants, but it is also appropriate to distribute the $4,200 NBCT salary increase across all applicants given that many probably would not have applied to the program without the salary inducement. Based on a 50 percent pass rate, this adds an additional $2,100 for a total per applicant cost of $4,400 for this program.[46] Given these figures and an assumed class size of 20, the estimated (lower-bound) cost per pupil of raising student achievement on reading tests by one standard deviation is about $7,300.[47] This is far more costly than estimates of some other popular educational interventions. For example, reductions in class size (from 25 to 20 students) are estimated to cost about $1,900 per student to achieve the same growth in reading scores (Levin and McEwan 2000).

We also wish to consider whether there are any benefits associated with identifying who the more effective teachers are. The total per NBCT identification cost (calculated above) is $8,800, and the most optimistic finding, which is about 20 percent of a standard deviation in math for third graders, yields an estimated cost-effectiveness ratio of $2,200 associated with NBCT identification.[48] This may be a far more cost-effective way of identifying teacher quality than paying a premium for teachers who hold a master's degree, since the evidence on the value of a master's degree is quite mixed (Hanushek 1986, 1997; Goldhaber and Brewer 2000; Greenwald et al. 1996). The benefit to students of identifying NBCTs depends on whether the identification (or the compensation associated with it) alters teachers' career paths. If, for instance, it does not cause them to remain in the teaching profession longer, there would be no

direct benefit from identifying which teachers were more effective, although there may be various indirect benefits.[49] We are unaware of any published empirical findings on whether becoming NBPTS certified alters a teacher's career path.

Conclusions

The significant interest and recent 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. Until now there has been no large-scale quantitative study to assess whether NBPTS has been successful in its goal to identify highly effective teachers. Thus, it has been unknown what the return has been thus far on the hundreds of millions of dollars that have been invested in NBPTS.

Many of our findings appear to confirm that the NBPTS assessment process is successful in identifying the more effective teachers among applicants to the program; the results are robust to a number of different model specifications, including those that account for the potential non-random matching of students to schools or classrooms. However, there are mixed findings on whether certification should be used as a signal of teacher quality. While we consistently find that teachers who will eventually be NBPTS certified are more effective, there are mixed findings about their effectiveness after being identified as an NBCT.

We find some evidence that going through the NBPTS assessment process adds to teachers' human capital, thus providing some support for the investment in the NBPTS program. Nevertheless, the question of whether the investment in NBPTS is worthwhile also depends on the impact of NBPTS certification on teachers' career paths. If NBCTs are, in fact, better teachers, the payoff inherent in their identification as such is certainly higher if they remain in

the teaching profession longer as a consequence of having been identified. Furthermore, the magnitudes of estimated NBCT impacts depend on the grade level and the type of students being taught, suggesting the value of identifying NBCTs may depend, at least in part, on their teaching assignments. These teacher career path issues certainly merit further empirical study.

Whether NBPTS certification serves as a signal of teacher quality is influenced both by the assessment process and the quality of applicants to the program. Our data encompass the early years of certification by NBPTS; it would not be surprising if the early cohorts of teacher applicants to the program differed markedly from those teachers who apply once the program has become better established. With this in mind, it will be important to determine if the estimated effects of NBCTs vary as the program matures and the applicant pool changes.

References

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

Ballou, D., and M. Podgursky. 1998. "The Case Against Teacher Certification." *The Public Interest* 132: 17–29.

Bond, L., T. Smith, W. Baker, and J. Hattie. 2000. *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.

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

Clotfelter C.T., H. F. Ladd, and J.L. Vigdor. 2003. "Teacher Sorting, Teacher Shopping, and the Assessment of Teacher Effectiveness." Paper presented at the American Association of Public Policy and Management, November.

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

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

Ferguson, R.1998. “Teachers’ Perceptions and Expectations and the Black-White Test Score Gap.” In *The Black-White Test Score Gap,* edited by C. Jencks and M. Phillips (273–317). Washington, D.C.: Brookings Institution Press.

Ferguson, R., and H. Ladd. 1996. “How and Why Money Matters: An Analysis of Alabama Schools.” In *Holding Schools Accountable*, edited by Helen Ladd. Washington, D.C.: Brookings Institute Press.

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

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

Goldhaber, D., and E. Anthony. 2003. *Teacher Quality and Student Achievement*. (ERIC Clearinghouse on Urban Education. New York: Teachers College, Institute for Urban and Minority Education).

Goldhaber, D., and D.J. Brewer. 1996. “Evaluating the Effect of Teacher Degree Level on Educational Performance.” In *Developments in School Finance 1996,* edited by J. William Fowler (197–210). Washington, D.C.: National Center for Education Statistics.

Goldhaber, D., and D.J. Brewer. 2000. “Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement.” *Educational Evaluation and Policy Analysis* 22: 129–145.

Goldhaber, D., and L. Cramer. 2003. “A Descriptive Analysis of the Distribution of NBPTS-Certified Teachers in North Carolina.” Paper presented at the twenty-fifth annual Association of Public Policy and Management (APPAM) Research Conference, Washington, D.C., Nov. 6–8.

Goldhaber, D., D. J. Brewer, and D. Anderson. 1999. “A Three-Way Error Components Analysis of Educational Productivity.” *Education Economics* 7(3): 199–208.

Goldhaber, D., D. Perry, and E. Anthony. 2004. “National Board Certification: Who Applies and What Factors are Associated with Success?” The Urban Institute Education Policy Center working paper. Washington, D.C.: Urban Institute.

Greenwald, R., L. Hedges, and R. Laine. 1996. “The Effect of School Resources on Student Achievement.” *Review of Education Research* 66(3): 361–396.

Hanushek, E. A. 1986. “The Economics of Schooling: Production and Efficiency in Public Schools.” *Journal of Economic Literature* 24(3): 1141–78.

Hanushek, E. A. 1992. “The Trade-Off Between Child Quantity and Quality.” *Journal of Political Economy* 100(1): 84–117.

Hanushek, E. A. 1997. “Assessing the Effects of School Resources on Student Performance: An Update.” *Educational Evaluation and Policy Analysis* 19(2): 141–64. (EJ 550 073).

Hanushek, E. A., J. F. Kain, and S.G. Rivkin. 1999. “Do Higher Salaries Buy Better Teachers?” National Bureau of Economic Research Working Paper No. 7082. Cambridge, Mass.: National Bureau of Economic Research.

Hanushek, E. A., J. F. Kain, and S.G. Rivkin. “Why Public Schools Lose Teachers.” *Journal of Human Resources* (forthcoming).

Hedges, L.V., R. Laine, and R. Greenwald. 1994. “A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes.” *Educational Researcher* 23(3): 5–14.

Lankford, H., S. Loeb, and J. Wyckoff. 2002. "Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis." *Educational Evaluation and Policy Analysis* 24(1): 37–62.

Latham, A. S., D. Gitomer, and R. Ziomek. 1999. “What the Tests Tell Us About New Teachers.” *Educational Leadership* 56(8): 23–26.

Levin, H. M, and P.J. McEwan. 2000. *Cost-Effectiveness Data,* 2d ed. Thousand Oaks, Calif.: Sage Publications.

Levinson, A. 1988. “Reexamining Teacher Preferences and Compensating Wages.” *Economics of Education Review* 7(3): 357–364.

Podgursky, M. 2001. "Defrocking the National Board: Will the Imprimatur of ‘Board Certification’ Professionalize Teaching?" *Education Matters*, 1(2): 79–82.

Stone, J.E. 2002. “The Value-Added Achievement Gains of NBPTS-Certified Teachers in Tennessee: A Brief Report.” College of Education, East Tennessee State University, 2:5.

Wilcox, D. D. 1999. “The National Board of Professional Teacher Standards: Can it Live up to its Promise?” In *Better Teachers, Better Schools*, edited by Marci Kanstoroom and Chester E. Finn, Jr. Washington, D.C.: The Thomas B. Fordham Foundation.

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

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

Variable	Non-applicants		Applicants, 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)
Fraction of free/reduced -price lunch students	.36 (.19)	.36 (.19)	.42 (.22)	.42 (.22)	.31 (.18)	.31 (.18)
Fraction of minority students	.36 (.25)	.36 (.25)	.42 (.38)	.42 (.28)	.32 (.22)	.32 (.22)
District/Community Characteristics						
Percent in community with at least a B.A.	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, teacher was an unsuccessful Current Applicant that year, teacher was a successful Current Applicant that year.
[b] Schools of excellence have 90 to 100 percent of their students performing at or above grade level.
[c] Schools of distinction have 80 to 89 percent of their students performing at or above grade level.[a] Free and reduced-price lunch information is only available at the student level starting in 1999.

Table 2. Selected Coefficients for Growth in Test Score Models
(Standard errors in parenthesis)

COLUMN	1	2	3	4	5	6	7	8	9	10	11	12
MODEL	**Reading Achievement Growth**						**Math Achievement Growth**					
Future Applicant				.16*** (.04)						-.21** (.10)		
Current Applicant			-.33*** (.09)	-.10 (.07)	-.36*** (.09)	-.57*** (.17)			-1.00*** (.10)	-.40*** (.07)	-1.06*** (.10)	-1.64*** (.19)
Past Applicant				.08 (.09)						.21*** (.04)		
Current NBCT	.19*** (.08)	.17** (.08)					.09 (.09)	.05 (.09)				
Future NBCT	.37*** (.05)	.36*** (.05)	.39*** (.05)		.35*** (.05)	.47*** (.09)	.62*** (.05)	.60*** (.05)	.69*** (.05)		.63*** (.06)	.67*** (.11)
New NBCT			.48*** (.13)		.49*** (.13)	.09 (.25)			1.11*** (.14)		1.18*** (.14)	1.60*** (.28)
Past NBCT			.23 (.14)		.28* (.15)	.08 (.27)			-.12 (.16)		-.09 (.16)	-1.17*** (.30)
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	.06	.06	.06	.06	.07	.60	.10	.10	.10	.10	.11	.57
Sample Size	609,160						611,517					

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

Note: 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-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.; 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)

STUDENT SUBGROUP	Free or Reduced-Price Lunch Recipients[a]		Non-Recipients of Free or Reduced-Price Lunch[a]		White		Black		Other	
COLUMN	1	2	3	4	5	6	7	8	9	10
MODEL	Read	Math	Read	Math	Read	Math	Read	Math	Read	Math
Current Applicant	-.32* (.19)	-.84*** (.20)	-.20 (.17)	-.68*** (.18)	-.21 (.11)	-.72*** (.12)	-.39** (.16)	-1.09*** (.18)	-.51 (.37)	-2.35*** (.41)
Future NBCT	.55*** (.19)	1.09*** (.21)	.32*** (.13)	.54*** (.14)	.34*** (.06)	.65*** (.06)	.31*** (.10)	.56*** (.12)	.85*** (.24)	.97*** (.27)
New NBCT	.98*** (.29)	1.50*** (.31)	.18 (.22)	.80*** (.23)	.37*** (.15)	.90*** (.17)	.32 (.25)	.74*** (.28)	1.36*** (.56)	3.09*** (.62)
Past NBCT	.60** (.30)	-.01 (.33)	.10 (.20)	.08 (.21)	.35** (.17)	.00 (.18)	.02 (.30)	-.23 (.33)	-.39 (.61)	-1.75*** (.68)
R^2	0.06	0.03	0.15	0.14	0.09	0.15	0.03	0.02	0.04	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.

Note: 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-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. 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-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.; 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 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)

GRADE LEVEL	Third Grade		Fourth Grade		Fifth Grade	
COLUMN	1	2	3	4	5	6
MODEL	Read	Math	Read	Math	Read	Math
Current Applicant	-.37** (.16)	-1.07*** (.17)	-.26* (.16)	-1.18*** (.18)	-.43*** (.13)	-.65*** (.15)
Future NBCT	.49*** (.09)	.96*** (.09)	.26*** (.08)	.35*** (.10)	.30*** (.08)	.52*** (.09)
New NBCT	1.01*** (.23)	1.73*** (.24)	-.09 (.22)	.96*** (.26)	.11 (.19)	.36* (.22)
Past NBCT	.63** (.31)	.09*** (.02)	.28 (.23)	-.80*** (.27)	.05 (.20)	-.44** (.22)
R^2	.05	.06	.02	.02	.02	.01
Sample Size	228,654	229,623	191,853	192,606	188,653	189,288

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

Note: 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-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.; 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."

Appendix A

Variable Definitions

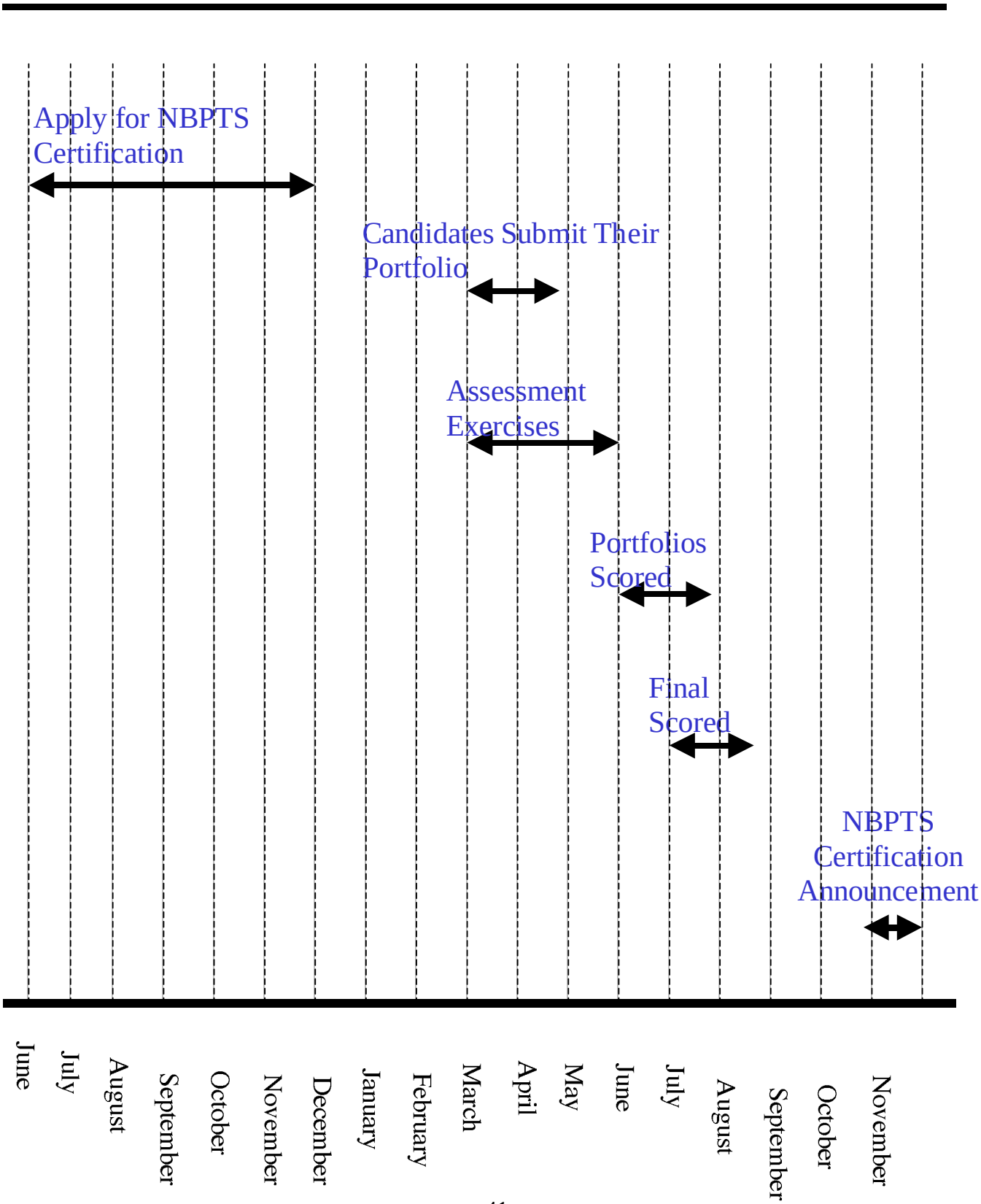
STUDENT VARIABLES	DATA SOURCE	VARIABLE DEFINITIONS
Post-test	NCDPI	Student end-of-year test score
Pre-test	NCDPI	Student test score from end of previous year, unless if student was a third grader. Third graders take a beginning-of-the-year test as pre-test.
Growth in test score in one year	NCDPI	Post-test 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 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 a master's degree
Ph.D./Other advanced degree	NCDPI	Teachers highest degree is a Ph.D. or other advanced degree
Past Applicant	NBPTS[a]	Teacher applied to NBPTS in a prior year
Current Applicant	NBPTS[a]	Teacher applies to NBPTS in current year
Future Applicant	NBPTS[a]	Teacher has not yet applied to NBPTS but we know applies in a future year for which we have data
Past NBCT	NBPTS[a]	Teacher had become an NBCT in a past year
Current NBCT	NBPTS[a]	Teacher is currently a NBCT
New NBCT	NBPTS[a]	Teacher becomes a NBCT in present year
Future NBCT	NBPTS[a]	Teacher becomes a NBCT in future year
SCHOOL VARIABLES	**DATA SOURCE**	**VARIABLE DEFINITION**
School of "excellence"	NCDPI web site	Schools of excellence have 90 to 100 percent of their students performing at or above grade level, determined by state standards.
School of "distinction"	NCDPI web site	Schools of distinction have 80 to 89 percent 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-lunch students	Common Core of Data	The fraction of students eligible for free-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 pre-kindergarten through twelfth 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 pre-school to grade twelve 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 B.A. 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 B.A.	Common Core of Data	Percent of residents with a bachelor's degree or higher degree in the 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]Note: For NBPTS variables, data is only available for teachers who applied for National Board Certification

Figure 1. NBPTS Application Process Timeline



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 with end-of-grade test scores (posts)	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 with valid pre- and posts 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

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 three-fold 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] This estimate was calculated using the total number of NBCTs as of November 2003 and the average pass rate for current applicants nationwide, which is 48 percent according to the Educational Testing Service (ETS) and NBPTS officials.

[3] This is calculated based on the product of the total number of current applicants, estimated to be 65,000 over the life of NBPTS, and the $2,300 assessment fee.

[4] Through September 2002, NBPTS has been appropriated federal funds of $119.3 million, representing approximately 45 percent of the more than $250 million in total money received for the National Board Certification project (http://www.nbpts.org).

[5] As of May 2003, 32 states and many localities offer at least one type of financial incentive (bonus or salary supplement) for teachers to become NBPTS certified, and some type of formalized support is offered in 49 states and approximately 486 local school districts, including the District of Columbia (http://www.nbpts.org/).

[6] See http://www.teachnow.la/common/incentives.php for more information on this NBPTS incentive.

[7] 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 credentials and student outcomes, and Hedges, Laine, and Greenwald (1994) or Card and

Krueger (1996), who find more positive results. We discuss "teacher effectiveness" in this paper in terms of the teachers' contributions toward student gains in achievement.

[9] Education productivity studies typically measure the size of the relationship between various quantifiable education factors and student achievement. Goldhaber, Brewer, and Anderson (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] Prior to 2001 (a time period corresponding to our data, described below), current applicants were judged based on the submission of a teaching portfolio containing two samples of student work, two teaching lessons, documentation of involvement in the parent community and in the professional community, and the completion of four assessment exercises on pedagogical and content knowledge related to the NBPTS certification area for which they were applying. Each portfolio component is scored and weighted according to a rubric specific to the certification area (as of August 2003, there are 27 NBPTS certification areas from general to social studies, science, English, art, math, etc. that are offered at various age levels, http://www.nbpts.org/candidates/guide/2_certif.html); in general the samples of student work and teaching lessons are weighted most heavily, followed by the assessment exercises and then documentation of community and professional involvement. In 2001, some of the NBPTS assessment procedures changed. For instance, teachers now complete only four portfolio exercises, all of which must involve samples of student work and self-reflective commentary by the teacher. The demonstration of community and professional involvement has been collapsed

into one portfolio entry rather than two separate ones. There are two additional assessment exercises, which focus more on testing content knowledge rather than pedagogical knowledge.

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

[12] 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, Gitomer, and Ziomek 1999).

[13] The Bond et al. study includes a total of 65 teachers, and the Stone study includes 16 teachers.

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

[15] We have information on free or reduced-price lunch status for 1999 only, when the state started collecting it.

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

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

[18] For most of the years of data in our study, over one-quarter of the nation’s NBCTs taught in North Carolina.

[19] 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 records 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 five districts serving the largest metropolitan areas in the state, until we received ten districts' responses on the question. All ten 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.

[20] 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.

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

[22] 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.

[23] 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).

[24] 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.

[25] For more information on how various growth/gain goals and measures are determined by the state, see http://newdev.www.ncpublicschools.org/accountability/reporting/sasr/2003 (July 28, 2003).

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

[27]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.

[28] 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.2** for the distribution of these students and teachers across years and grades.

[29] 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.

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

[31] All else equal, in reading and math, growth in test score for one year 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; non-participants 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.

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

[33] 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 larger than our sample of 303 Current NBCT teacher observations.

[34] The decrease in the magnitudes of Current NBCT and Future NBCT is not statistically significant in either the reading or math models.

[35] See discussion of the model that is the exception to this (column 4 of **table 2**) in the next section, which further explores the effect of NBPTS application.

[36] There are only 87 teacher observations in which a teacher had been NBPTS certified in the past.

[37] We have 213 unique past NBPTS applicant teacher observations in reading and math, and 1,345 and 1,349 unique future NBPTS applicant teacher observations in reading and math, respectively.

[38] For instance, all past applicants are teachers who went through the NBPTS assessment process in 1997–98 or an earlier year, while future applicants are teachers who went through the NBPTS assessment process in 1997–98 or a later year.

[39] Although we do not report them here, the results of these models are available from the authors upon request.

[40] We also estimated teacher fixed effects models (without Past NBCT due to too little variation in this variable) to examine the possibility that unobservable variables at the teacher level might be correlated with teachers' NBPTS status. Our math findings were consistent with our other findings, in that the coefficient on Current Applicant is negative and significant and that the coefficients on New and Future NBCT are positive and significant, while all of the coefficients on the NBPTS variables (Current Applicant, New NBCT, Future NBCT) are in the same direction as previous reading models but are now insignificant. These results should be interpreted very cautiously, however, as 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.

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

[42] Furthermore, the sum of Current Applicant and New NBCT is positive and statistically significant in the third grade reading and math models, implying that having a NBCT *in the year of application* is beneficial for students in that grade.

[43] This figure is calculated distributing across NBCTs the cost of assessing unsuccessful applicants and adding this to the $6,500 direct cost.

[44] It also depends on the subject area, but in most elementary schools NBCTs are teaching both math and reading.

[45] The .4 test score point figure is based on the differential between the coefficients on Future Applicant and Past Applicant in the math achievement model (column 10 of **table 2**).

[46] Again, this is actually a lower-bound estimate of the cost since the salary supplement is good for the 10-year life of the certificate.

[47] This is calculated by dividing the $4,400 per applicant cost by 20 students to obtain a cost per student of $220, then dividing this by the estimated achievement effect, .03 standard deviations.

[48] This is calculated by taking the per student cost of identifying a NBCT, $8,800 divided by 20 (the assumed class size), and dividing that by the average assumed gain (.2 standard deviations) per student.

[49] For example, NBCTs may model effective teaching techniques for other teachers or bring about other positive changes to schools.