National Board Certification as Professional Development: An Empirical Study of the National Board for Professional Teaching Standards Process

Final Report

Linda Cavalluzzo, Ph.D.
Principal Investigator
CNA

Lisa Barrow, Ph.D.
Co-Principal Investigator
Federal Reserve Bank of Chicago

Stephen Henderson, Ed.D.
Co-Principal Investigator
Briarwood Enterprises, LLC

with

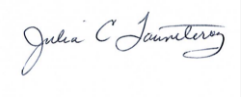
Christine Mokher, Ph.D.
Thomas Geraghty, Ph.D.
CNA

Lauren Sartain
University of Chicago

IRM-2015-U-009975
April 15, 2014

CNA
ANALYSIS & SOLUTIONS

Approved for distribution:



Julia C. Fauntleroy
Managing Director, CNA Education

This document contains the best opinion of CNA at the time of issue.
It does not necessarily represent the opinion of the Sponsor.

Distribution Unlimited

Contents

Executive summary

The National Board for Professional Teaching Standards (NBPTS) is a professional organization that provides national certification to teachers who apply for and meet the Board's standards of performance for "accomplished" educators. The certification process is voluntary, and it is a time-consuming and rigorous one, requiring applicants to furnish a portfolio containing videos of their instruction, copies of their students' work, and written reflections on their instruction, as well as to complete online exercises that assess their pedagogical and subject-matter knowledge.

The National Board certification (NBC) process is a research-based program that was developed over a 10-year period with financing from the National Science Foundation, the U.S. Department of Education, and private funders. Only experienced, certified educators with at least a bachelor's degree are eligible to apply. The certification process can take a few months to two years. Teachers who are unsuccessful may refine and resubmit portions of their application and/or retake the exercises to raise their score and achieve certification on a second or third attempt.

Because of the significant resources involved, both in the development of the standards and in the application process, there has been a good deal of attention focused on NBC's value and effectiveness.

In 2008, the National Academy of Sciences' National Research Council released a report reviewing the evidence on the NBC's effectiveness. The Council concluded that, "The evidence is clear National Board certification distinguishes more effective teachers from less effective teachers with respect to student achievement" (2008, p. 179). But the extant literature left understudied, and unresolved, whether the certification process itself also improves teachers' effectiveness by augmenting human capital—the intrinsic capability of a teacher to teach effectively, which may be increased through experience, education, training, and professional development.

The Council also noted that the large-scale statistical studies pertaining to National Board certification focused almost exclusively on teachers in Florida and North Carolina, and on the elementary grades. Furthermore, virtually all of the analyses focused only on the test scores of students in mathematics or reading.

Study goals

This study responds to a request from the NBPTS to analyze National Board certification among high school teachers in understudied subject areas and locales to help fill gaps in the research literature. We also were asked to use multiple indicators of performance.

Approach

The research team selected two new locales for this analysis, the Commonwealth of Kentucky and the Chicago public schools. Chicago, a racially and ethnically diverse city with a population of more than 2.8 million, has one of the largest urban school districts in the country. Kentucky, by contrast, is a largely rural state with some suburban and urban areas, including the Louisville/Jefferson County metro area, population 750,000. Together, these two locales encompass a full range of public school settings.

The proliferation of longitudinal data systems that allow researchers to link students to their subject-area teachers and to track student performance over time provides new opportunities to examine NBPTS processes in these new locations. In addition, both school systems use ACT's Educational Planning and Assessment System (EPAS) to monitor the academic progress of their high school students. EPAS comprises three assessments: the EXPLORE®, given in grade 8 or 9; the PLAN®, given in grade 10; and the ACT®, given in grade 11 or 12. Each assessment includes subtests in English, mathematics, science, reading, and writing. In this study we use test scores in the first three subject areas to examine outcomes for high school students whose teachers participated in the NBC process and for high school students whose teachers did not participate.

In addition to examining student test scores, we also conducted classroom observations of the instructional practices of high school teachers in science and mathematics, comparing a sample of NBC applicants and similar teachers not pursuing this certification. We

conducted these observations at baseline—that is, in the semester when the NBC applicants first submitted their applications for certification—and then again in each of the next two semesters. Most of the comparison teachers came from the same schools as the NBC applicants and were observed on the same days.

We used the Leadership by Design (LBD) classroom observation instrument to assess instruction. Teachers were rated on nine different dimensions of instruction: *lesson overview, instructional overview, questioning, classroom atmosphere, concept development, teacher's content knowledge, learning climate, classroom management,* and *assessments.* Teachers also were given an overall instructional-quality rating by the site observers.

Research questions

In order to get a thorough understanding of the effects of National Board certification, we addressed the following four questions:

1. Does the National Board certification process influence teachers' classroom practices?

As measured by student test scores:

2. Are National Board–certified teachers more effective than other teachers?

3. Are applicants who attain National Board certification more effective than applicants who do not?

4. What effect, if any, does the National Board certification *process* have on teacher effectiveness?

Question 1 is addressed by examining instructional practices over time for NBC applicants compared with non-applicants. To address questions 2–4, we compare outcomes for students taught by National Board–certified teachers with those taught by non-certified teachers, developing three different modeling frameworks that measure, respectively, the efficacy of National Board certification in "signaling" teacher effectiveness, "screening" for teacher effectiveness, and "human capital" formation that increases teacher effectiveness.

Findings

Ratings of the instructional practices of NBC applicants exceeded those of non-applicants at baseline on six of the nine teaching-quality subscales, as well as on the overall rating of instructional quality. However, there was little evidence of growth in instructional quality over the observation period for either applicants or non-applicants.

Our analyses of student test scores considered five different model specifications, and student achievement gains were estimated for PLAN and ACT scores in English, mathematics, and science. The baseline model controls for a rich set of student characteristics, including prior test score. Subsequent models add school characteristics and the average pretest score of all students assigned to a given teacher. These models help to correct for the nonrandom assignment of students to schools and to teachers that may affect measurements of teacher effectiveness. A final model replaces school characteristics with school fixed effects, providing comparisons of teacher effectiveness within schools.

We found evidence that Board certification is an effective "signal" of teacher quality. Although effect sizes varied, these results generally held across locales, test types, and subject areas. The estimated effect sizes are similar to those found elsewhere in the literature, and are smallest when National Board–certified teachers (NBCTs) were compared with other teachers in the same schools.

The "screening" models compared student outcomes based on the amount of instruction students had from teachers who ever earned or would later earn Board certification during the study period and the amount of instruction students had from teachers who applied for National Board certification but were never certified. These models found some evidence that NBC effectively screens applicants. Results were strongest for mathematics, and weakest for English, and generally did not reveal differences for within-school comparisons.

We were unable to find evidence of a "human capital" effect indicating that teacher effectiveness increased over time, based on student test scores for teachers in our sample, including those who advanced through the NBC process from pre-applicant to applicant or from applicant to post-applicant.

Conclusions and recommendations

Using data for high school teachers and their students from Chicago and Kentucky public schools, we found evidence that National Board certification is an effective signal of teacher quality, based on student test scores. We also found some evidence that the certification process successfully screens applicants based on their effectiveness. But we were unable to find evidence that the certification process itself enhances the instructional quality or effectiveness of teachers who choose to go through it.

Our analysis of the professional development value of the National Board certification process as measured by changes in instructional practices was limited by the length of time over which we were able to observe teachers' practices for changes, as well as by the inability to identify and observe teachers prior to their joining the applicant pool. It is quite likely that new applicants have already spent time prior to formally applying for Board certification reflecting on their practices, and possibly taking steps to improve those practices. Indeed, programs such as NBPTS's own Take One! are designed to help teachers prepare for the application process before formally applying. Our inability to observe teachers before they formally file their application may cause our estimates to understate the true impact of NBC on teaching practices.

The analysis of improvements in teachers' effectiveness as measured by their students' test scores also was limited by the four-year period of the provided data, which dictated the number of teachers we were able to observe in each stage of the certification process.

It is important to keep in mind that our findings about the human capital effects only pertain to the experienced teachers eligible to apply for National Board certification. The results shed no light on the potential of the certification process to improve the instructional practices of less-experienced teachers (i.e., with fewer than three years of teaching) who are not eligible, or of less-able teachers who do not apply for certification.

Nor does our analysis examine the role that the certification process might play in helping to identify specific areas of improvement for teachers who go through the process, or identify which elements of

the applicant portfolio are most closely linked to teacher effectiveness, as measured by student test scores.

Given that the National Board certification process has repeatedly demonstrated the ability to distinguish between more and less effective teachers, school systems should think about how to make good use of this tool. For example, school systems could use National Board certification as a gatekeeper for advancement or as part of the tenure decision process, where tenure decisions are implemented at a later point in the teaching career path than the criteria most school systems currently use for those decisions.

Introduction

One of the most important issues facing education policy-makers is how to prepare students to be productive citizens in an increasingly competitive global economy. Evidence from state and national assessments provides a mixed picture as to whether states are successfully doing so. While state accountability systems suggest that the proportion of students meeting state benchmarks is rising, performance on the National Assessment of Educational Progress (NAEP) has been relatively stagnant, especially in mathematics and among 17-year-olds (Rampey, Dion, & Donahue, 2009).[1]

The teacher quality literature suggests that teachers are the single most important school-based input into student learning, and that teacher quality (as measured by a teacher's contribution to student achievement on standardized tests) varies considerably across schools and also within a single school (Aaronson, Barrow, & Sander, 2007; Goldhaber, 2002; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004). These measures of teacher quality are, however, largely unrelated to any of the teacher characteristics generally available, such as highest level of education (Clotfelter, Ladd, & Vigdor, 2007; Goldhaber, 2007); years of teaching experience beyond the first two or three (Clotfelter et al., 2007; Goldhaber, 2002; Rivkin et al., 2005); or indicators of ability such as selectivity of undergraduate institution or test scores (Goldhaber, 2002; 2007; Harris & Sass, 2007; Kane, Rockoff, & Staiger, 2008). So teachers are important to the learning process, but it is difficult to pinpoint specific measures that identify high-quality teachers.

Improving teacher quality has been central to significant national education initiatives in the Bush and Obama administrations. No Child Left Behind (NCLB) is national legislation passed in 2001 that

1. The NAEP is the only nationally representative assessment of student achievement in the United States. It is funded by the U.S. Department of Education. Samples of 4th, 8th, and 12th grade students take the NAEP every other year.

increased emphasis on state accountability systems. One of NCLB's major mandates was that all students should be taught by "high-quality" teachers. The definition of high quality was that all teachers must be fully certified, have at least a bachelor's degree, and demonstrate content area knowledge—although research (cited above) published since the passage of NCLB indicates that these particular indicators are not necessarily markers of high-quality teachers.

In 2009–10, the U.S. Department of Education initiated a grant competition called Race to the Top, in which states compete for federal education funding. To be competitive for these grants, states have to show commitment to improving the quality of teaching by designing and implementing better teacher evaluation systems, increasing equitable access of students to good teachers and good principals, and improving the state of teacher preparation programs and teacher support. The component of teacher and principal quality gets the most weight in the competition.

One way teachers can demonstrate their skill level and successes in the classroom is by earning certification from the National Board for Professional Teaching Standards (NBPTS). National Board was established to help professionalize the field of teaching by providing an accepted definition of what "accomplished" teaching is and recognizing teachers who do their jobs exceptionally well. An original goal of National Board certification (NBC) was to build an authentic assessment system that could reliably measure what experienced teachers should know and be able to do (Carnegie Task Force on Teaching as a Profession, 1986). Educators would volunteer to participate in the program and those who successfully demonstrated the appropriate level of professionalism and expertise would be awarded a nationally recognized certificate attesting to that level of demonstrated performance.

Since being established in 1987, NBPTS has certified more than 100,000 teachers, and countless more have participated in the application process (NBPTS, 2013). Large investments have been made in the development of the National Board certification program. As of September 2005, the National Science Foundation and the U.S. Department of Education had appropriated more than $149 million dollars to it, and nongovernment funders had spent an additional $261 million (Cohen & Rice, 2005). Applicants for certification (or

more typically, their sponsoring school systems) also incur substantial costs. As a result, there is a great deal of interest in identifying and measuring the full value to education systems of encouraging teachers to become National Board certified.

This study uses a two-pronged approach to examine the effectiveness of National Board–certified teachers and NBC applicants. As described in the first part of this report, we use classroom observations of teachers in the state of Kentucky and in Chicago Public Schools (CPS) to examine the quality of instructional practices of National Board applicants and non-applicants and whether teachers' instructional practices change over time. We observe outcomes for National Board certification applicants at the beginning, middle, and end of the process, and compare the results with a control group of non-applicants.

As described in the second part of this report, we analyze administrative data for teachers and students, again from Kentucky and Chicago Public Schools, matching students to their demographic characteristics, multiple years of standardized test scores, and teachers. This allows us to examine signaling and screening effects of National Board certification, as well as human capital formation—that is, any professional development benefits of the NBC process, as measured by improvement in test scores of the students of National Board–certified teachers.

Through this analysis we want to better understand how the National Board certification process relates to teaching effectiveness and to changes in teaching practice, and thus to improvements in student learning. Specifically we seek to answer these questions:

1. Does the National Board certification process influence teachers' classroom practices?

As measured by student test scores:

2. Are National Board–certified teachers more effective than other teachers?

3. Are applicants who attain National Board certification more effective than applicants who do not?

4. What effect, if any, does the National Board certification *process* have on teacher effectiveness?

This report begins by describing the role of National Board in improving student learning and by reviewing the relevant literature. Second, we describe the setting, the data sources, and the characteristics of the schools, teachers, and students in our sample. Third, we describe our methods and findings from the classroom observations; we also describe the methods and findings from our analyses of student test scores. We conclude by summarizing the key findings, the limitations of this study, and the implications both for future research and for practice.

The role of National Board in improving student learning

NBPTS developed a rigorous, multifaceted evaluation program for the purpose of identifying highly effective ("accomplished") teachers. Applicants can select from among 25 certificate areas, which are based on the age of the students taught and the subject area of instruction (not all subject areas are available in every age category).[2]

Table 1: National Board certification subject areas and age categories.

Subject areas	Age categories
Art	Early childhood (ages 3–8)
Career and technical education	Middle childhood (ages 7–12)
English as a new language	Early and middle childhood (ages 3–12)
English language arts	Early childhood through young adulthood (ages 3–18+)
Exceptional needs specialist	Early adolescence (ages 11–15)
Generalist	Adolescence and young adulthood (ages 14–18+)
Health education	Early adolescence through young adulthood (ages 11–18+)
Library media	
Literacy	
Mathematics	
Music	
Physical education	
School counseling	
Science	
Social studies–History	
World language	

To apply, teachers must assemble and submit a portfolio of specific materials, including artifacts from their classroom instruction and student work, video of their classroom interactions with students, written reflections analyzing the instructional practice evident in the videos and student work, and a written statement that demonstrates

2. For more information, see http://www.nbpts.org/certificate-areas.

their involvement in activities outside the classroom that benefit student learning. In addition, they must pass six in-depth computer-based "exercises," essentially assessments of their content and pedagogical knowledge in their specialty area (NBPTS, 2011).

In all, the process can take many months to two years. Applicants submit their application forms, fees, and proof of eligibility and begin developing their portfolios between February and December of the first year. Eligible applicants then take the computer-based assessments between March and June of the second year. At least one portfolio entry must be submitted by May of year two. Applicants have a maximum of two years to complete all the requirements, with the caveat that no portfolio entry can be more than 12 months old. Applicants do not find out their certification status until the following November–December.

In an evolution to the original process, teachers who do not pass all sections of the certification may reapply and resubmit materials for the section(s) they did not pass previously. The reapplication cycle is 1 year, as opposed to the initial 2-year application window. Once awarded, National Board certification is valid for 10 years, at which point teachers must reapply if they are interested in maintaining their certification status.

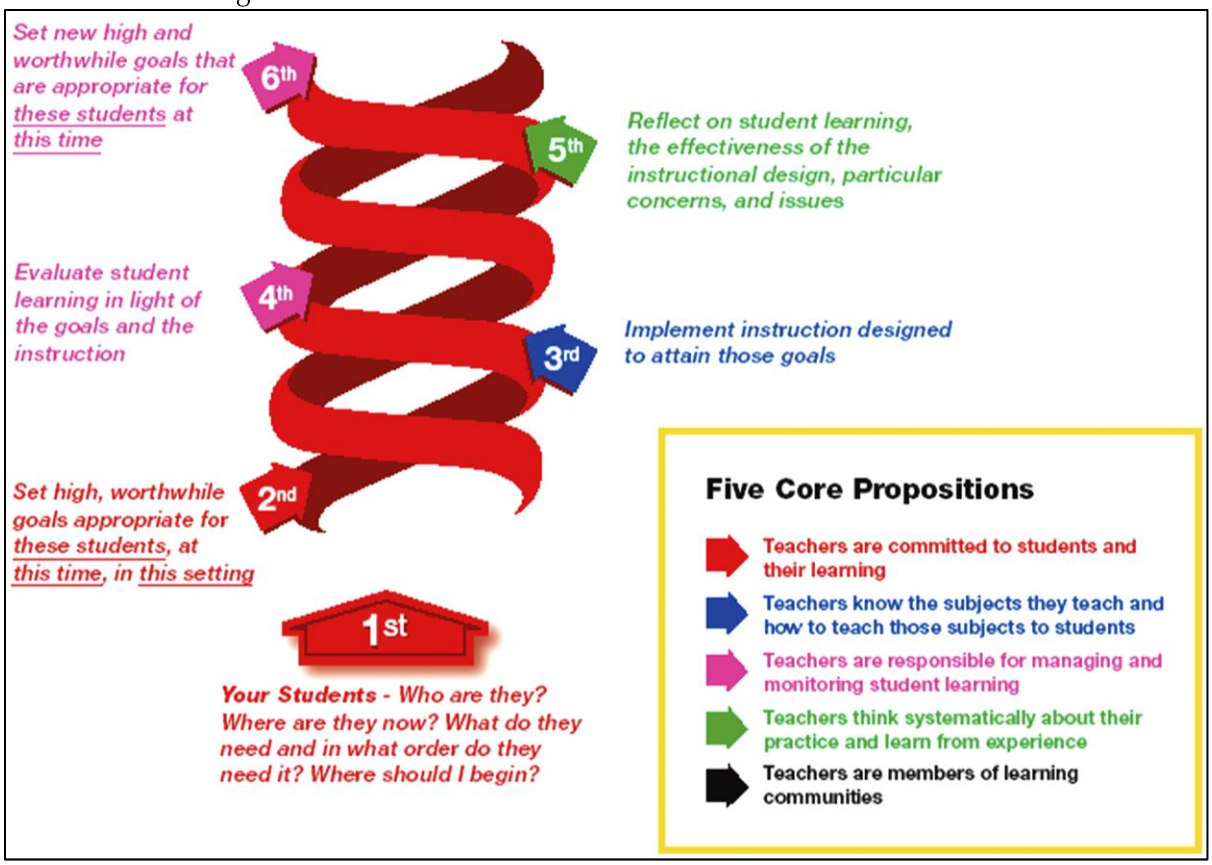
The National Board certification process defines "accomplished" teaching based on five core propositions (NBPTS, 2002):

- Proposition 1: Teachers are committed to students and their learning.
- Proposition 2: Teachers know the subjects they teach and how to teach those subjects to students.
- Proposition 3: Teachers are responsible for managing and monitoring student learning.
- Proposition 4: Teachers think systematically about their practice and learn from experience.
- Proposition 5: Teachers are members of learning communities.

NBPTS uses an "Architecture of Accomplished Teaching Helix" to illustrate what accomplished teaching looks like (see Figure 1). The process begins with the teacher understanding the needs of the stu-

dents and setting appropriate goals for them. Then the teacher implements instruction based on those goals, evaluates learning related to the goals, and reflects on students' learning. This is a continuous process, in that the teacher continually repeats it by setting new goals that are appropriate for students at the current time.

Figure 1: National Board for Professional Teaching Standards "Architecture of Accomplished Teaching Helix."



SOURCE: NBPTS, 2012.

Each NBC applicant is expected to demonstrate the five core propositions in their video recording of a whole class discussion, commentary on the instruction evident in the video, and responses to written questions that guide the teacher to address the certification standards and the core propositions. The written commentary is expected to be analytic and reflective, demonstrating the teacher's understanding of his or her own teaching practices and the students' learning.

Teachers who decide to apply for National Board certification generally have many support options available to them. Many teachers ask a colleague to help them reflect on their practices and build a strong portfolio. A preparatory professional development program offered by NBPTS called Take One! will provide teachers with information about the certification standards and allows them to submit a video portfolio entry for scoring prior to formally applying. Some districts and state departments of education, including the Kentucky Department of Education (KDE) and the Chicago Public Schools (CPS), have central office staff members dedicated to helping teachers become National Board certified.

In Chicago, teachers have at least two options (one through the district and another through the teachers' union) for ongoing candidate support during the National Board application process. These programs provide weekly or biweekly meetings for candidate teachers to come together to review and revise their portfolios, as well as counseling on whether or not the time-consuming process is a good fit for them. In Kentucky, the Kentucky Education Association offers professional learning opportunities for teachers interested in applying for certification or renewal. It also provides training for educators who are interested in serving as mentors to National Board candidates. Further, many postsecondary schools of education offer programs to help teachers prepare for the rigors of National Board certification.

Putting all the pieces together, completing the NBC process requires a significant investment of time and effort. Because only teachers with at least three years of teaching experience are eligible to apply, National Board certification does not help principals make hiring decisions with less-experienced teachers. Yet, simply identifying high-quality teachers has no direct effect on the number of them in the profession. What impact, then, can National Board certification have on student learning?

In this study, we investigate the main ways in which National Board can improve the quality of classroom teaching. The first has been the subject of much academic research—that being National Board certified can serve as an indicator of teacher quality. This implies both that high-quality teachers apply for National Board certification (the signaling effect) and that the NBC process does a good job of screening applicants and awarding certification to the most qualified (the

screening effect). If certification is a good indicator of teacher quality, then principals and district administrators can use National Board certification to inform their staffing and leadership decisions with experienced teachers. Namely, given a large enough supply of National Board–certified teachers, principals and school districts can improve average teacher quality by staffing a large number of teaching positions with National Board–certified teachers.

A second way in which National Board certification might improve average teacher quality is by using the process as part of a framework for better managing the teacher workforce. If National Board certification were part of a deliberate system aimed at improving the overall quality of instruction, if it were used, for example, as part of a revised tenure, compensation, or advancement system, more able candidates might choose to enter, or stay, in teaching.

A third way in which National Board certification might improve average teacher quality is by changing and improving teachers' practices. In other words, perhaps the NBC process itself contributes in terms of "human capital" by developing better teachers, regardless of the outcome of their applications.

We discuss, in turn, each of these roles: the role of National Board certification as a signal to identify high-quality teachers; the ability of the NBC process to screen less-effective applicants from more-effective applicants; and the human capital role of the NBC process itself in improving instructional quality through teacher professional development.

Identifying high-quality teachers

The end goal of most education policy interventions is to improve student outcomes, and the main mechanism for increasing student learning is to ensure that students are exposed to high-quality teaching. One strategy is to replace underperforming teachers with higher-quality teachers. While this approach might at first glance seem simple to implement, there are many complicating issues. First and foremost, researchers and policy-makers continue to grapple with how to measure teaching effectiveness, since the observable teacher characteristics available in most datasets have little correlation with measures of student learning.

As an alternative to using traditional teacher characteristics such as years of experience and highest level of education completed, National Board certification could be used by teachers to signal that they are high quality. If so, principals could use this information to allocate resources and staff more effectively. Perhaps within a school, principals might give National Board teachers more desirable assignments in order to keep them in a school; principals in other schools might try to single out National Board–certified teachers in the hiring process; and so on. In short, certified teachers might have more flexibility both in their current positions and in the larger teacher labor market.

There is evidence that obtaining National Board certification has signaling value—that teachers with National Board certification are indeed of higher quality than teachers who are not certified (Cantrell, Fullerton, Kane, & Staiger, 2008; Cavalluzzo, 2004; Clotfelter, Ladd, & Vigdor, 2007; Goldhaber & Anthony, 2007). Most studies that identify the signaling effect of National Board certification compare certified teachers (NBCTs) and noncertified teachers, making statistical adjustments to account for the fact that teachers who participate in certification might be different from those who do not.

These effect sizes are generally statistically significant, though small. For example, McCaffrey and Rivkin (2007) found that compared with other, noncertified teachers in the state, North Carolina NBCTs raised 4th and 5th grade math scores on the state-mandated accountability test by 7 to 8 percent of a standard deviation, and reading scores for the same grades by 4 to 5 percent of a standard deviation. They further found that in Florida, NBCTs raised 4th and 5th grade reading scores by 2 to 4 percent of a standard deviation compared with noncertified teachers; Florida NBCTs had no statistically significant effect, however, on 4th and 5th grade math scores. These results are broadly consistent with those of several other studies (Clotfelder et al., 2007; Goldhaber & Anthony, 2007; Harris & Sass, 2006; Sanders, Ashton, & Wright, 2005). All of these studies find modest effects in reading, but the results are more mixed in math.

Research suggests, too, that the NBC process is a good screening mechanism for identifying high-quality teachers. The screening effect refers to the ability of the National Board certification process to distinguish more-effective from less-effective teachers who apply for cer-

tification. As such, National Board–certified teachers are more effective than are applicants who complete the application process but do not achieve certification, as measured by student achievement (Cavalluzzo, 2004; Clotfelter et al., 2007; Goldhaber & Anthony, 2007; Sanders et al., 2005). In general, these studies find that students taught by National Board–certified teachers make statistically significantly larger test score gains than those taught by teachers who applied but were not certified. Effect sizes tend to be larger for math than for reading (Hakel, Koenig, & Elliott, 2008).

The literature cited here focuses almost exclusively on statistical comparisons in just two states, Florida and North Carolina, and on elementary school students. In this study, we expand on the existing literature—providing evidence from two additional locations, Kentucky and Chicago. We also focus exclusively on high school teachers.

Human capital development

In the context of education, "human capital" can be defined as the intrinsic capability of a teacher to engage in effective instruction. A teacher's human capital stock can be increased through investment in education, training, and professional development activities (Eide and Showalter, 2010). As with any educational intervention, the quality of professional development varies, from good to bad and everything in between. Research on professional development in Chicago Public Schools suggests that teachers benefit most from training that promotes ambitious, intellectually challenging instruction; occurs frequently and over time; exposes the teacher to content in his or her subject area; and features developments in pedagogical techniques (Smylie, Allensworth, Greenberg, Harris, & Luppescu, 2001). The U.S. Department of Education defines high-quality professional development as sustained and content focused, aligned with state learning standards, and focused on developing understanding of "scientifically proven" instructional techniques (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007).

Overall, the literature shows little to no effect of most professional development programs on student outcomes (e.g., Harris & Sass, 2007; Jacob & Lefgren, 2004; Podgursky, Springer, & Hutton, 2010). In particular, much of the funding for professional development is spent on "one-shot" workshops or other events not shown to translate

into improvements in student outcomes (Garet, Porter, Desimone, Birman, & Yoon, 2011).

There is some research, however, identifying characteristics of high-quality professional development programs (e.g., Jacob & Lefgren, 2004), and the National Board certification process appears to have many of these. The NBC application process itself is sustained over time, and the application materials include a portfolio of lessons, assessments, and reflections prepared by the teacher and based on the students in his or her actual classroom. Although the original motivation for establishing NBPTS was not to build a strong professional development program, it is clear that its certification process has the markings of one. As a result, it is reasonable to expect that participation in the NBC process could improve a teacher's instruction, and that better instruction would translate into better student outcomes.

Here, the question we are interested in answering has to do with the third way in which National Board certification can improve student learning—that is, does participation in the NBC process itself improve that teacher's effectiveness, regardless of whether or not the applicant completes it and/or achieves certification? Is the NBC process effective professional development?

The extant literature leaves understudied, and unresolved, whether National Board certification is more than a good signal of and screen for identifying high-quality teachers. Many studies that try to capture its human capital effects compare teachers who are at different stages in the certification process (before applying, applying, and after applying). They typically find that teachers' effectiveness declines marginally while they are applying, which could be a result of their spending so much time and energy on their portfolio that it distracts from their teaching (Clotfelder et al., 2006; 2007; Goldhaber & Anthony, 2007; Harris & Sass, 2006; McCaffrey & Rivkin, 2007). These same studies produce mixed results about gains in teacher effectiveness after the application process ends.

It is worth noting that there are limitations in the current research. Any observable gains in student learning might simply be due to certified teachers being better able to signal and sort into schools or to getting different teaching assignments after being certified. Gains

could just be a function of certified teachers now teaching higher achieving students or in higher achieving schools.

We propose a different approach to estimating the human capital effects: comparing individual teachers against themselves over time using a teacher fixed effects model. Although this approach has had limited use in the research literature (e.g., Harris & Sass, 2006), it should result in more accurate estimates of the ability of the National Board certification process to increase teacher human capital.

Changing classroom practices

While the majority of research on the effects of National Board certification has relied on administrative datasets (i.e., test scores), several studies have looked at the effect of the process on teachers' classroom practices, including instruction and classroom management. Darling-Hammond, Atkin, Sato, Chung, Dean, and Greenwald (2007) used teacher-submitted lesson videotapes and student work samples, interviews, and surveys to assess the effects of the certification process on high school math and science teachers. This study randomly assigned teachers to two groups—one group who applied for National Board certification, and a second group who postponed their application until after the study. The study's attrition rate was high: about 75 percent of the teachers in the initial sample dropped out, leaving a final sample of only 16 teachers. The study found some evidence that teachers who went through NBC improved their formative assessment practices more than did nonparticipants: teachers who applied for certification were found to use a wider variety of assessment methods and better integrated assessment with instruction.

Other studies have used survey evidence to assess the self-reported views of teachers who have gone through National Board certification (Indiana Professional Standards Board, 2002; Yankelovich Partners, 2001). Typically, the surveys are conducted only after teachers complete the certification process, so there is no way to disentangle whether differences in practices were preexisting or due to participation (Hakel et al., 2008). Nevertheless, teachers tend to report NBC helped them improve their teaching and increased their ability to reflect on their teaching practices and incorporate the results of this reflective activity into their instruction.

We will provide further evidence of the effect of the National Board certification process on classroom practices through a series of classroom observations. Our observations of National Board applicants are conducted at three points in time: once as teachers begin the certification process for the first time, once in the middle of the process, and once at the end. Observations are also conducted at similar times for a set of control teachers not participating in certification.

These observations provide additional support in testing whether the National Board certification process is an effective screening or signaling mechanism, and whether it is effective professional development. For example, the screening effect would be supported if National Board applicants start out with higher ratings on their classroom observations than do non-applicants. This would indicate that teachers who self-select into the certification process tend to be higher-quality teachers to begin with. The human capital hypothesis would be supported if NBC applicants demonstrate greater gains in instructional quality over time than do non-applicants. This would suggest that participating teachers may be learning new information through certification that is improving their teaching.

Description of the data

The setting

The data we analyzed for this study (both the classroom observations of teacher instruction and the student test scores) are from public school across the state of Kentucky and from the Chicago Public Schools district. Kentucky is an ideal state for this study. First, National Board enjoys strong support there. Through the efforts of teachers and the financial support of the Teachers' National Certification Incentive Trust Fund, the state has become one of the largest producers of NBCTs: 1,116 or about 4 percent of the teaching workforce.[3] This compares favorably with the national average of about 2 percent. To our knowledge, however, there has been no notable research on the effectiveness of NBCTs compared with noncertified teachers in the state.

Kentucky has other appealing features, as well. It is largely rural, yet has suburban and urban centers, including the Louisville/Jefferson County metro area, with a 2010 population of about 750,000.[4] Furthermore, Kentucky uses ACT's nationally recognized Educational Planning and Assessment System (EPAS) to monitor growth in student achievement over time. The state also has a longitudinal data system that uses unique identifiers to track students across the state and over time. The data system links students to their teachers, to the courses they enroll in, and to their statewide assessments.

Chicago was selected as a second location to broaden the research base of the study. The city of Chicago has a population of 2.8 million, and its very large urban school system is home to 1,158 NBCTs, or 36 percent of all NBCTs in the state of Illinois. Like other large urban districts, CPS is racially and ethnically diverse. Further CPS has been using EPAS since 2003 and has the results stored in a longitudinal da-

3. Calculated based on data provided by NBPTS.

4. Data from the U.S. Census (www.census.gov).

ta system, permitting development of study results that are comparable to those in Kentucky.

Data sources

Our analysis of student outcomes relies on administrative data from all CPS high schools and all public middle and high schools in the state of Kentucky. Student-level data files were provided by CPS through the University of Chicago Consortium on Chicago School Research, and the Kentucky Department of Education, respectively. These data files include school enrollment records, course records linked to the teacher of record for the course, test scores, and student demographic characteristics. In both locations, we have four years of data, allowing us to measure changes in student outcomes over time for three cohorts of students for each analysis. In Kentucky, the data are available for school years (SYs) 2007/08 through 2010/11; in Chicago, the data are available for 2008/09 through 2011/12.

Student test scores

Both CPS and Kentucky use EPAS, which consists of three tests: EXPLORE®, PLAN®, and ACT®. The EXPLORE is administered in the fall of grade 8 in Kentucky and the fall of grade 9 in CPS. In both locations, the PLAN is administered in the fall of grade 10; and the ACT is administered in the spring of grade 11.

According to ACT, Inc., the tests are aligned so that the score of the next test in the series can be predicted based on the prior test. Each test results in five sub-area scores: English, mathematics, reading, science, and writing. The composite score is the average of all of the sub-area scores except for writing. EPAS also has the advantage of being nationally normed, so we know how student performance compares with other students in Illinois, for example, or around the country.

We conduct two sets of analyses for this study: the first uses the EXPLORE as a pretest and the PLAN score as the outcome measure; the second analysis uses the PLAN as a pretest and the ACT score as the outcome. The analysis sample includes only students who have both pretest and posttest scores. The majority of students took each test one time; however, if a student has more than one test score, we

use the score from the date of the earliest test, so the results are comparable to students who took the test only once.

We standardized the scale scores for each test by subtracting the national mean score on the corresponding test from the student's test score, and then dividing by the national standard deviation. This allows the magnitude of the effects to be directly compared across subjects, test (EXPLORE, PLAN, ACT), and locales (CPS, Kentucky). Results are examined separately for English, math, and science. We also examine the results for the three subjects combined.[5]

Student information

Both CPS and Kentucky administrative data collected on students include basic demographic information, such as gender and race/ethnicity, as well as socioeconomic status (based on free/reduced-price lunch (FRL) eligibility and special education status (students with Individualized Education Programs (IEPs)). Date of birth was used to calculate each student's age at the beginning of each school year. In addition, Kentucky has an indicator for English as a Second Language (ESL) status, and the number of days the student was absent during the school year.

The analytic sample in Chicago includes 69,741 students for the PLAN analysis and 48,546 for the ACT analysis. In Kentucky, the sample sizes are 80,490 for the PLAN and 114,465 for the ACT. (Some 34,903 Kentucky students are in both the PLAN and the ACT samples.)

Teacher information

NBPTS provided certification application data for teachers in Chicago Public Schools and Kentucky starting with the 2000 applicant cohort and ending with the 2012 applicant cohort. These data include application date(s), number of times applied, and the outcome of each application for teachers of all subjects and grade levels. We also have information about the subject area and age category for certification.

[5] We did not examine test scores in reading or writing because those topics do not align to a specific teacher.

Over this 13-year period, there were 4,658 unique applicants from CPS, and 44 percent of them achieved National Board certification. From Kentucky there were 4,746 unique applicants, and 54 percent of them achieved National Board certification. Most applicants applied one time (71 percent for CPS, 67 percent for Kentucky); only about 1 percent of teachers applied more than three times.

There is no unique teacher ID number in the data file from NBPTS that can be used to merge the file with the teacher records in the administrative data files from CPS or KDE. Instead, we matched the records using teachers' first names, last names, and email addresses. We started by identifying any exact matches in either address or first and last name in both files. Then we looked for cases where the names were similar but not exact. We manually checked these records and compared other characteristics in the two files, such as school name and subject area, to determine whether the records appeared to belong to the same person. The match rate could be expected to be less than 100 percent because our administrative data files include only public school teachers, while the file from NBPTS includes other applicants such as administrators and private school teachers. For the years of our analysis, the match rate is 83 percent in Kentucky and 78 percent in Chicago.

In Chicago, the National Board data could be linked to the CPS personnel data, giving us access to a richer set of teacher covariates. The personnel data include characteristics such as number of years teaching in the district, level of education, area of teacher certification, and demographic attributes. Similar data are not available for teachers in Kentucky.

In order to link students to their teachers, we also used transcript files that account for all the courses in which a student enrolls and the teachers of each course. In Chicago, each course in the transcript file was coded as "core" (English, mathematics, or science—to map to the EPAS sub-area test scores) or "non-core." For this analysis we restrict the dataset to core courses. Core courses all count toward the Illinois state graduation requirements. In Kentucky, we coded courses as English, math, or science, based on standardized state course codes. We also reviewed course descriptions provided by KDE and coded courses as primary or elective based on these descriptions.

For both Chicago and Kentucky, we include only teachers of primary courses in the analysis. If the student took both a primary course and an elective course in a particular subject area, we included the record from the primary course in the analysis and included a dummy variable in the model to indicate that the student was also enrolled in an elective course in the same subject area. In Kentucky, we also coded whether the course level is unknown, basic (e.g., remedial courses), regular, or advanced (e.g., honors, Advanced Placement, and International Baccalaureate).

Students who have more than one primary course in the same subject area taught by more than one teacher were flagged as having multiple teachers. Conversely, students without any courses in the core subject area were flagged as having no teachers. While we cannot identify the individual teacher responsible for teaching these students in those particular semesters/years, we do not want to drop them from the analytic dataset. (See Appendix D for more information on construction of the analytic file.)[6]

School information

Most of the school-level data we use for Kentucky come from the Common Core of Data housed at the U.S. Department of Education's National Center for Education Statistics. The Common Core of Data makes publicly available characteristics about each school across the country, and the data can be aggregated up to the district, state, or national level. Covariates include school size, student-teacher ratio, student-administrator ratio (district level), percent Black students,

6. For the Kentucky ACT sample (114,465 students): in math, 3.2 percent of students attended a block-scheduled course, 9.0 percent had multiple teachers, and 5.6 percent had no teacher (could not be matched). For English, the percentages were 3.6 percent block, 6.7 percent multiple, and 5.0 percent missing. For science, they were 3.8 percent block, 23.3 percent multiple, and 9.8 percent missing. For the CPS PLAN sample (69,741 students), 12 percent of students had multiple teachers in math, while 1.5 percent did not have a designated teacher and 0.2 percent had no math class. For English, the percentages were 15 percent multiple, 1.6 percent missing, and 0.1 percent no English class; and for science, 7 percent multiple teachers, 1.4 percent missing, and 1.5 percent no science class. See Appendix D for additional information.

percent Hispanic students, percent FRL students, per pupil spending, and school locale. For CPS, we calculate school-level variables from the student-level data including averages of student neighborhood socioeconomic indices. We also use the EPAS data provided by CPS and KDE to calculate school-level average scores on ACT, PLAN, and EXPLORE in each subject area.

Characteristics of sample schools, teachers, and students

The percentage of students in the sample who ever had an NBCT during the timeframe of the analysis is 11 percent in Kentucky and 28 percent in CPS. There are statistically significant differences between students who had a class with one or more NBCTs and students who did not on all of the characteristics we examined. As shown in Table 2, students who never had an NBCT had lower test scores on the pre-tests (EXPLORE and PLAN) in math, English, and science; and had higher rates of absences from school than students taught by an NBCT. Students who never had an NBCT were also less likely to be Black, Hispanic, or female and more likely to be categorized as FRL, IEP, or ESL. This indicates that the population of students taught by NBCTs differs from students taught by non-certified teachers.

Table 2: Comparison of student characteristics, by whether the student ever had a National Board–certified teacher.

	Kentucky			CPS		
	Had NBCT: No	Had NBCT: Yes	Difference	Had NBCT: No	Had NBCT: Yes	Difference
Average EXPLORE pretest score in math (PLAN sample)	14.7	15.7	-1.0*	14.5	17.0	-2.5*
Average EXPLORE pretest score in English (PLAN sample)	14.0	15.1	-1.1*	13.5	16.3	-2.8*
Average EXPLORE pretest score in science (PLAN sample)	16.1	16.9	-0.8*	15.7	17.7	-2.0*
Average PLAN pretest score in math (ACT sample)	16.8	18.2	-1.4*	15.1	17.7	-2.6*
Average PLAN pretest score in English (ACT sample)	16.1	17.3	-1.2*	14.4	16.9	-2.5*
Average PLAN pretest score in science (ACT sample)	17.8	18.7	-1.0*	16.2	18.0	-1.8*

	Kentucky			CPS		
	Had NBCT: No	Had NBCT: Yes	Difference	Had NBCT: No	Had NBCT: Yes	Difference
Average number of absences per year	10.2	8.3	1.9*	NA	NA	
% Black	8.6	13.8	-5.2*	43.7	31.5	12.2*
% Hispanic	2.1	3.2	-1.1*	44.0	43.8	0.2
% Female	50.0	51.3	-1.3*	51.8	55.4	-3.6*
% Free or reduced-price lunch	47.8	38.8	9.1*	69.6	63.0	6.7*
% Individualized Education Program	7.8	3.8	4.0*	12.3	5.1	7.3*
% English as a Second Language	2.6	4.3	-1.7*	NA	NA	

NOTES: N=160,052 students in Kentucky (34,903 in both the PLAN and ACT samples) and 89,002 students in CPS (29,285 in both the PLAN and ACT samples). Of Kentucky students, 16,853 had an NBCT in math, English, or science during the analysis timeframe. Of CPS students, 24,715 had an NBCT in math, English, or science. Significance was calculated using two-tailed t-tests of mean ratings for students who had an NBCT during the analysis timeframe compared with students who did not. *=difference is statistically significant at the .05 level.~=difference is statistically significant at the .1 level.

Approximately 5 percent of teachers in Kentucky and 17 percent of teachers in CPS in the sample ever applied for National Board certification during the timeframe of the analysis (see Table 3). Among those teachers who do apply in Kentucky, 52 percent achieve certification, 36 percent do not achieve, and 12 percent have unknown outcomes because they completed the certification process after the analysis period. In CPS, 48 percent of NBC applicants achieve certification, 21 percent do not achieve, and 31 percent are unknown because they withdrew from the process or completed after the last date reported from NBPTS.

Table 3: Number and percentage of teachers in the sample who ever applied for National Board certification and who achieved it during the timeframe of the analysis.

	Kentucky		CPS	
	N	%	N	%
Teacher ever applied for NBC?				
Yes	423	4.6	665	16.5
No	8,839	95.4	3,357	83.5
Teacher ever achieve NBC?				
Yes	221	52.3	321	48.3
No	153	36.2	138	20.8
Unknown	49	11.6	206	31.0

NOTE: N=9,262 teachers in Kentucky and 4,022 teachers in CPS.

The percentage of schools that had an NBCT during the analysis period was 64 percent in Kentucky; 84 percent of schools in the CPS sample had an NBCT during the analysis period. As shown in Table 4, Kentucky schools with NBCTs are more likely to be in suburban areas, and less likely to be in rural areas than school without NBCTs. Thus, it is not surprising that Kentucky schools with NBCTs have larger total enrollments than schools without NBCTs. Schools in Kentucky with NBCTs also have fewer FRL students and higher test scores on the EXPLORE and the PLAN than schools without NBCTs. Similarly, Chicago schools that had at least one NBCT are larger on average than schools without any NBCTs. Chicago schools with any NBCTs also have somewhat higher average test scores.

Table 4: Comparison of school characteristics, by whether the school ever had a National Board–certified teacher.

	Kentucky			CPS		
	Had NBCT: No	Had NBCT: Yes	Difference	Had NBCT: No	Had NBCT: Yes	Difference
Total enrollment	667.6	939.9	-272.3*	532.5	1134.6	-602.1*
Student-teacher ratio	17.4	18.2	-0.7	15.6	15.3	0.3
Student-administrator ratio (in district)	212.2	221.8	-9.6	NA	NA	
% Black students	12.2	13.1	-0.8	74.6	54.8	19.7
% Hispanic students	1.8	2.4	-0.6	18.9	35.0	-16.1
% Free or reduced-price lunch students	61.7	49.8	12.0*	92.2	85.3	6.9
Per pupil spending ($)	10,294	10,392	99*	NA	NA	
% Urban schools	13.3	18.6	-5.2	NA	NA	
% Suburban schools	7.8	18.0	-10.2*	NA	NA	
% Town schools	23.3	21.6	1.8	NA	NA	
% Rural schools	55.6	41.9	13.6*	NA	NA	
School-level average EXPLORE score in English	13.9	15.3	-1.5*	11.9	13.5	-1.6
School-level average EXPLORE score in math	14.5	16.1	-1.6*	12.3	14.1	-1.8*
School-level average EXPLORE score in science	16.4	17.2	-0.9*	14.3	15.5	-1.2*
School-level average PLAN score in English	12.1	13.2	-1.1*	12.8	14.4	-1.6
School-level average PLAN score in math	13.1	13.9	-0.8*	13.4	15.2	-1.8~
School-level average PLAN score in science	14.7	15.5	-0.8*	15.4	16.6	-1.2~

NOTES: N=359 schools in Kentucky and 100 schools in CPS. Significance was calculated using two-tailed t-tests of mean ratings for schools that had an NBCT during the analysis timeframe compared with schools that did not. *=difference is statistically significant at the .05 level. ~=difference is statistically significant at the .1 level.

This page intentionally left blank.

Classroom observations

One aspect of our evaluation involves classroom observations of a sample of NBC candidate teachers and a sample of other teachers with similar characteristics in similar classroom settings who are not pursuing certification. The goal of this part of the study is to chart and compare applicants and non-applicants and to examine these teachers' use of effective instructional practices over time.

Changes in instructional quality are examined for 27 math and science teachers in Kentucky and Chicago over a three-semester period. Where possible, we observed each teacher twice in the same semester to improve the quality of the data, using the average of the two observation scores for our analysis. However, it was not always possible to arrange for two observations each semester due to scheduling constraints.

We use the observations to address the following research question:

1. Does the NBPTS certification process influence teachers' classroom practices?

Comparing any gains in instructional quality for the two samples lets us draw conclusions about the effects of participation in certification. This study design requires a comprehensive observation instrument to document what is observed, a tool for assigning numeric scores to the instructional practices observed, and consistent and reliable data collection and scoring procedures to maintain the internal validity of these data.

Classroom observation instrument

We selected the Leadership by Design (LBD) classroom observation instrument for use in the study (see Appendix A).[7] This instrument

7. We slightly modified the instrument by moving the classroom context indicators from the front of the instrument to the end. This change was

has been widely used in Kentucky and elsewhere; classroom observation data have been collected using the LBD instrument for more than 3,000 teachers in more than 250 elementary, middle, and high schools in seven different states. Projects using the LBD include work funded by the U.S. Department of Education and the National Science Foundation. The LBD also has been adopted by the National Science Teachers Association as a program improvement tool to help assess and improve the quality of instruction in middle school and high school classrooms.

LBD measures the quality of instructional practices in science and math, as well as capturing information about the classroom setting. The instrument is completed during observations lasting 45 to 90 minutes by trained observers with subject-matter expertise. The rubric itself consists of 33 elements spanning nine dimensions: *lesson overview, instructional overview, questioning, classroom atmosphere, concept development, teacher's content knowledge, learning climate, classroom management,* and *assessments.*

The data collected through the LBD is descriptive in nature. Observers make notes, for example, about the types of questioning techniques used by the teacher, the amount of student investigation or research, the type of basic and higher-level skills being developed, and the teacher's use of formative and/or summative assessments to measure student learning. The LBD acts as a memory device for the observer; the data collected from the LBD are not used directly to rate the quality of instruction.

Rubric for scoring classroom observations

To assign numeric scores to the observation data collected with the LBD, we developed a "LBD Classroom Observation Rubric" for this study (see Appendix B). Prior to using the rubric in our evaluation, the research team piloted it using observations of a small sample of teachers (see Appendix C). The pilot test did not identify any problems in transferring the observation data to the rubric, and also con-

---

made so that the evaluator would not be distracted by the classroom context while evaluating teaching quality.

firmed that the scoring data produced by the rubric were internally consistent.

The rubric consists of nine instruction-related subscales, plus an overall rating. The subscales are based on the average rating of three to five specific items aligned with the LBD instrument. Each item on the rubric is scored on an integer scale of 1–5, with 5 being the highest rating and 1 the lowest. Scores of 3 and below show areas needing improvement. The rubric also has a subscale for the classroom's *physical setting,* collected to provide baseline contextual information and not used to evaluate the teacher or quality of instruction.

After rating each of the items on the rubric, observers assign an overall rating of the quality of the instruction. This overall rating takes into account the observer's overall impression, including the effectiveness of instruction, alignment with objectives and standards, student engagement, and instruction to develop students' higher-order thinking skills. Observers are required to write comments corresponding to the overall rating to provide context for understanding why the rating was selected. Table 5 provides an example of the rating rubric for the overall classroom observation rating.

Table 5: LBD Classroom Observation Rubric rating scale for overall classroom observation rating.

Rating	Description
5	Instruction was of high quality and effective for all students; evidence that instruction was based on clearly defined objectives that were fully aligned with standards; all students were engaged in activities requiring higher level thinking skills
4	Instruction was of high quality and effective for most students; evidence that instruction was based on clearly defined objectives that were aligned with standards; most students were engaged in activities that required higher level thinking skills.
3	Instruction was of good quality and effective for many students; instruction appeared to be based on student objects somewhat aligned to standards; some students had an opportunity for higher level thinking skills development.
2	Instruction was of mediocre quality and effective for only a small portion of the students; little evidence that instruction as based on student objectives; instruction had minimal impact on student learning.
1	Instruction was of poor quality and was not effective for any students; no evidence that instruction was based on student objectives; learning was not based on instruction provided.

Recruitment of teachers

Each year NBPTS provided us with contact information for any new NBC applicants in Chicago and Kentucky. All new applicants in math

and science at the high school level were contacted and asked to participate in the study. Teachers who agreed to participate committed to being observed twice per semester for three consecutive semesters. These semesters correspond to the beginning, middle, and end of the National Board certification cycle.

Even after repeated attempts, only about half of the teachers we contacted agreed to participate. Teachers have many competing demands on their time, particularly those engaging in a time-consuming endeavor such as National Board certification. In addition, many teachers we contacted expressed reluctance at having an unknown observer come into their classroom.

Once the NBC applicants were recruited for the classroom observations, the principal of each school was asked to identify a similar (control) teacher in the same school who was not an NBC applicant. The research team requested that the teachers selected for the control group be state certified in math or science and have at least three years of teaching experience (to match NBC eligibility requirements). We do not have any evidence, but we expect that the principal probably selected as the control someone who was perceived to be a "good teacher," so there could be no perception that the school was not doing a good job. We also expect that the control teachers were selected because they were confident and willing to have an outside observer in their classroom. This means that the control group may include higher-quality teachers than the "average" teacher. Four NBC applicants had no matched control teacher because the principal of their school declined to name one.

We were able to recruit 32 teachers, whom we observed at least once; 27 of these teachers were observed all three times. Due to the small number of new NBC applicants who agreed to participate in the study, we recruited over several semesters.

Observations were conducted from the spring semester of SY 2010/11 through the fall semester of SY 2013/14. Table 6 shows a total of 27 teachers were observed at all three time points: 9 in math and 18 in science. Fifteen (15) of the teachers were NBC applicants and 12 were not. The analysis includes only teachers observed at three time points, so that the sample is the same for the comparisons at each time point. Five additional teachers were observed only once

or twice (e.g., because the teacher retired or left the school); these were excluded from the analysis.

Table 6: Number of teachers observed at three time points, by location.

	Kentucky	Chicago	Total
Math	1	8	9
Science	10	8	18
NBC applicants	6	9	15
Non–NBC applicants	5	7	12
Total	11	16	27

Classroom observation process

The LBD and rubric data were collected during prearranged classroom visits by site observers. Observers were not informed which teachers were NBC applicants and which were not. The developer of LBD (Co-Principal Investigator Dr. Stephen Henderson) trained all observers annually to use the LBD and scoring rubric. All observers are experienced math or science teachers who also have used the LBD instrument for previous studies.

Participating teachers were instructed to teach the same lesson they would normally teach on the day of the visit and to use the same techniques/materials they would normally. During the classroom observation, the observer filled out the LBD instrument, marking items as they were observed. While in the classroom, the observer also looked for the following as used and available: text and other instructional resources currently being used; any student workbook(s) used; sample assessment given by the teacher; and a student laboratory manual or portfolio.

Following the observation, the teachers were asked to participate in a 5- to 10-minute debrief interview with the observer. Questions asked include the following:

- What were the goals of today's class?
- What went well in this class? What didn't go well?
- What are your thoughts on goals for tomorrow's class?

After the site visit, the observer reflected on the observation and, using the completed LBD instrument, filled out our LBD Classroom Observation Rubric. The classroom materials and the discussion with the teacher also enabled the observer to better understand what was observed, facilitating more accurate completion of the rubric.

Completed LBD observation instruments and scoring rubrics were collected by Dr. Henderson from the classroom observers following their classroom visits. Copies of the completed data collection instruments were provided to CNA for independent analysis.

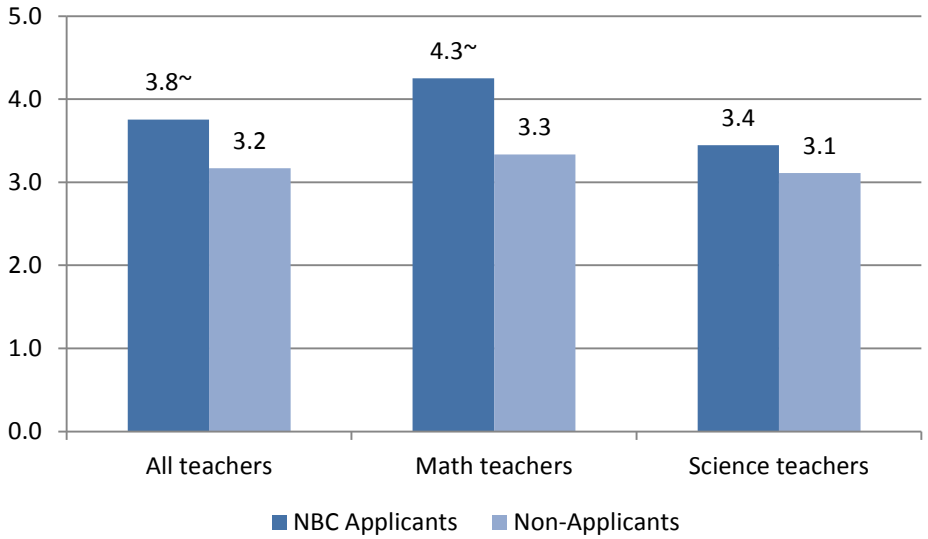
Results: Baseline ratings for NBC applicants and non-applicants

We begin our discussion of the results by describing the baseline (initial) observations for all teachers, comparing NBC applicants and non-applicants. This section examines whether National Board applicants have higher ratings of instructional quality than non-applicants just as the former start the steps in the certification process.

The mean difference between NBC applicant and non-applicant scores was calculated, and statistical significance was tested using a two-tailed t-test for unequal sample sizes and unequal variances. Figure 2 shows the average overall rating scores for all NBC applicants and non-applicants, as well as for math and science teachers separately.

There is some evidence that for all teachers, the overall ratings are higher for National Board applicants than for non-applicants (mean of 3.8 versus 3.2; a difference of 0.6, $p<.10$). However, the difference was statistically significant only for math teachers, with those NBC applicants' average overall rating (4.3) being a full point higher than the non-applicants' overall rating (3.3).

Figure 2: Average overall ratings for the baseline observations for NBC applicants and non-applicants, overall and by subject.



NOTES: Scale ranges from 1 (low) to 5 (high). N=27 teachers. Significance was calculated using two-tailed t-tests of mean ratings for NBC applicants compared with non-applicants. *=difference is statistically significant using a 95 percent confidence level. ~=difference is statistically significant using a 90 percent confidence level.

Next, we compared the baseline observations for NBC applicants and non-applicants on each of the nine rubric subscales. Table 7 shows that there is substantial variation in the range of scores for both NBC applicants and non-applicants. For both groups, the minimum scores for most subscales are below 3.0, and all of the maximum scores are between 4.5 and 5.0. The standard deviations range from 0.74 to 1.21 on a 5-point scale.

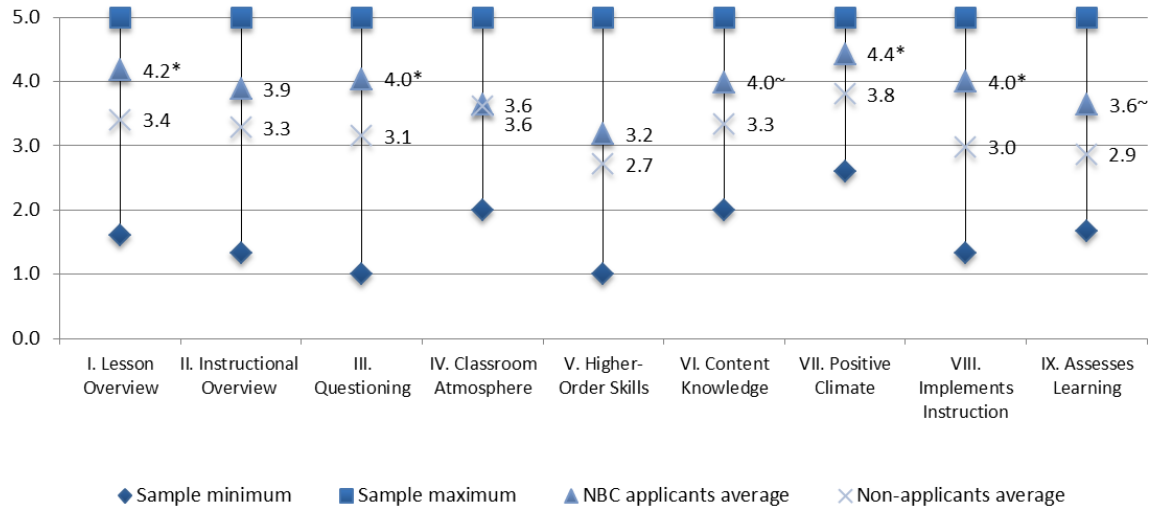
Table 7: Descriptive statistics for baseline observation scores for NBC applicants and non-applicants for each of the nine subscales and the overall rating scale on the LBD Classroom Observation Rubric.

	NBC applicants				Non-applicants			
	Min.	Max.	Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.
Overall rating	3.00	5.00	3.75	1.03	2.00	5.00	3.17	0.86
Lesson overview	3.20	5.00	4.17	0.81	1.60	4.60	3.40	0.86
Instructional overview	2.83	5.00	3.88	0.89	1.33	5.00	3.28	0.89
Questioning	2.84	5.00	4.04	0.97	1.00	4.75	3.15	1.04
Classroom atmosphere	2.10	4.50	3.63	0.78	2.00	5.00	3.61	0.81
Higher-order skills	1.50	5.00	3.18	1.20	1.00	5.00	2.71	1.19
Content knowledge	2.50	5.00	3.98	0.87	2.00	5.00	3.33	0.96
Positive climate	2.60	5.00	4.43	1.20	2.60	4.60	3.80	0.74
Implements instruction	2.67	5.00	3.99	0.87	1.33	5.00	2.97	1.21
Assesses learning	1.67	5.00	3.63	0.77	1.67	4.67	2.86	1.04

NOTE: Scale ranges from 1 (low) to 5 (high). N=27 teachers.

As shown in Figure 3, the average score for NBC applicants is statistically significantly higher than the average score for non-applicants on six of the nine rubric subscales: *lesson overview, questioning, content knowledge, positive climate, implements instruction,* and *assesses learning.* Variation in scores was greatest for the *questioning* and *higher-order skills* subscales, which ranged along the full scale of the rubric, with a 4-point difference between the minimum (1) and the maximum (5) scores.

Figure 3: Range of baseline observation scores for NBC applicants and non-applicants for each of the nine subscales on the LBD Classroom Observation Rubric.



NOTES: Scale ranges from 1 (low) to 5 (high). N=27 teachers. Significance was calculated using two-tailed t-tests of mean ratings for NBC applicants compared with non-applicants. *=difference is statistically significant using a 95 percent confidence level. ~=difference is statistically significant using a 90 percent confidence level.

Below we describe the subscales for which there was a statistically significant difference between NBC applicants and non-applicants at baseline. We also provide an example of an observer's description from a geometry class observation of an NBC applicant in math who had a high rating (4.5 or above) on each of these subscales.

- **Lesson overview:** NBC applicant mean=4.2 versus non-applicant mean=3.4, a difference of 0.8 ($p<.05$). This rating takes into account communication of lesson objectives, use of instructional resources to achieve the objectives, presentation of content in an accurate and grade-level-appropriate manner, place of the lesson in the instructional sequence, and choice of seating arrangements for the lesson. In the observation for the sample teacher's class, the observer commented,

  > The lesson on finding patterns on a unit circle was completely explored through pre-assessment, hands-on investigation, printed charts and diagrams, and technology. Students were seated in functioning groups for both individual and group accountability.

- **Questioning:** NBC applicant mean=4.0 versus non-applicant mean=3.1, a difference of 0.9 ($p<.05$). This rating takes into account the quality of the questions, student participation in questioning, use of strategic or target-centered questions for formative assessment, and feedback to students on responses. In the observation for the sample teacher's class, the observer commented,

  > Questions were purposeful and designed to discover misconceptions. All students were expected to be accountable in answering questions either in whole group discussions or in small groups. Wait time was not particularly intentional, but the type of questions required students to reason, and feedback was qualitative.

- **Content knowledge:** NBC applicant mean=4.0 versus non-applicant mean=3.3, a difference of 0.7 ($p<.10$). This rating includes communicating content knowledge to students, connecting content to life experiences, using instructional strategies appropriate for content, and guiding students to understand lesson content from various perspectives. The observer noted,

  > The teacher is exceptional and is able to orchestrate the various stages of the lesson seemingly effortlessly. He made a couple of realistic connections with the clock (unit circle and degrees) and periodic behavior (the sine curve). Students considered patterns on the unit circle chart, diagram of circle, sine curve using coordinate plane, string, and spaghetti, and on graphing calculator.

- **Positive climate:** NBC applicant mean=4.4 versus non-applicant mean=3.8, a difference of 0.6 ($p<.05$). To achieve a high rating, teachers must communicate high expectations, establish a positive learning environment, value and support student diversity, foster mutual respect between teacher and students and among students, and provide a safe environment for learning. The observer commented,

  > Students knew they were expected to accomplish tasks in assigned periods of time, and activities changed often to meet the needs of all students. The teacher had incredibly good rapport with students.

- **Implements instruction:** NBC applicant mean=4.0 versus non-applicant mean=3.0, a difference of 1.0 ($p<.05$). To achieve a high rating, teachers must implement instruction based on student needs and assessment data, use resources effectively, and manage instruction to facilitate higher-order thinking. The observer commented,

  > As the teacher monitored groups, he asked questions to determine if clarification was needed or if students were ready to explain their new pattern to the whole group, or figure out their misconceptions. While students did the warm-up, the teacher took roll and spot-checked every student's homework, and collected garbage and materials as students did an assessment at the end of class. Each student in small groups had a task to accomplish. Students had ample purposeful independent and group processing and reflection time.

- **Assesses learning:** NBC applicant mean=3.6 versus non-applicant mean=2.9, a difference of 0.7 ($p<.10$). This rating includes using assessments aligned with learning objectives, using a variety of formative and summative assessments to measure learning, and adapting assessments to accommodate diverse learning needs. The observer noted,

  > Besides the warm-up and homework checks, students took a 3-minute pre-assessment on their knowledge of the unit circle and then checked it themselves with the chart, answered questions asked by the teacher throughout the activities and by other students (teacher directed others to answer questions), reported on their patterns discovered, and demonstrated their learning with a writing assignment at the end (choice of 2 prompts—explain the pattern on the calculator or explain the concept in a short paragraph).

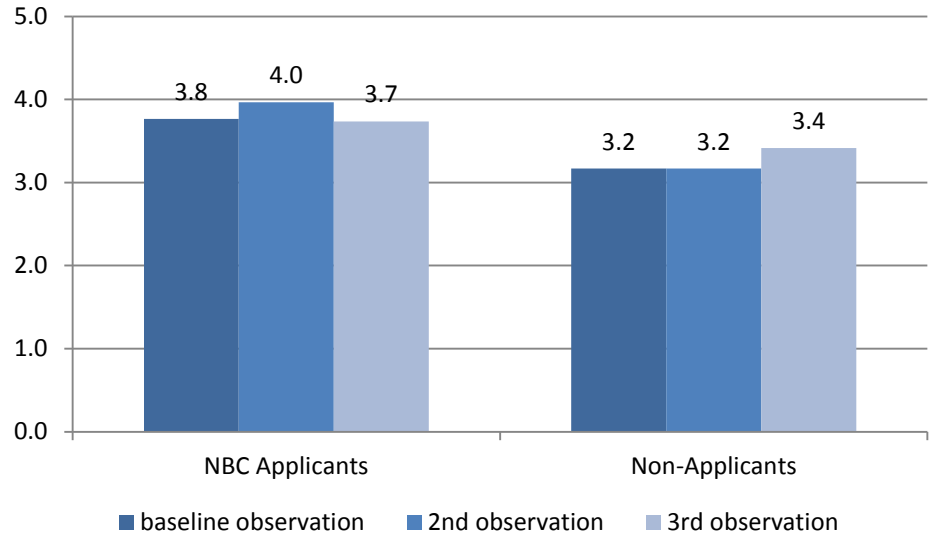
Results: Change in ratings over time for NBC applicants and non-applicants

To examine the effects of the National Board certification process, we compared the ratings from the baseline observations with the subsequent revisit observations, to see how the teachers' ratings change over time. Figure 4 shows the average overall ratings on the baseline,

second, and third observations for NBC applicants and non-applicants.

There are minimal differences between the baseline and subsequent observations for both groups of teachers, and none of the differences is statistically significant. This suggests that undergoing National Board certification does not have a distinguishable effect on teachers' overall quality of instruction.

Figure 4: Average overall ratings over time for NBC applicants and non-applicants.



NOTES: Scale ranges from 1 (low) to 5 (high). N=27 teachers. Significance was calculated using two-tailed t-tests of mean ratings for the baseline observation compared with each subsequent observation. *=difference is statistically significant using a 95 percent confidence level. ~=difference is statistically significant using a 90 percent confidence level.

We do not necessarily expect the National Board certification process to significantly affect the teachers' classroom practices on all of the LBD subscales, which is why we examine the differences separately for each subscale. It is also possible that certain subscales may be affected at different points in the application process, or that teachers' timing of implementing certain instructional elements may vary. Thus, we conduct comparisons both between the NBC applicants' second observation and baseline observations and between the NBC applicants' third observations and baseline observations. We also check whether there are any changes over time in the ratings for the non-applicants, although we do not anticipate significant differences

since these teachers are working under business-as-usual circumstances.

For the non-NBC teachers, we find no statistically significant differences between their scores at baseline and the second time or third time they were observed on any of the nine LBD rubric subscales (see Table 8). For NBC applicants, only one of the subscales (*classroom atmosphere*) has a statistically significantly increase over the baseline observation.

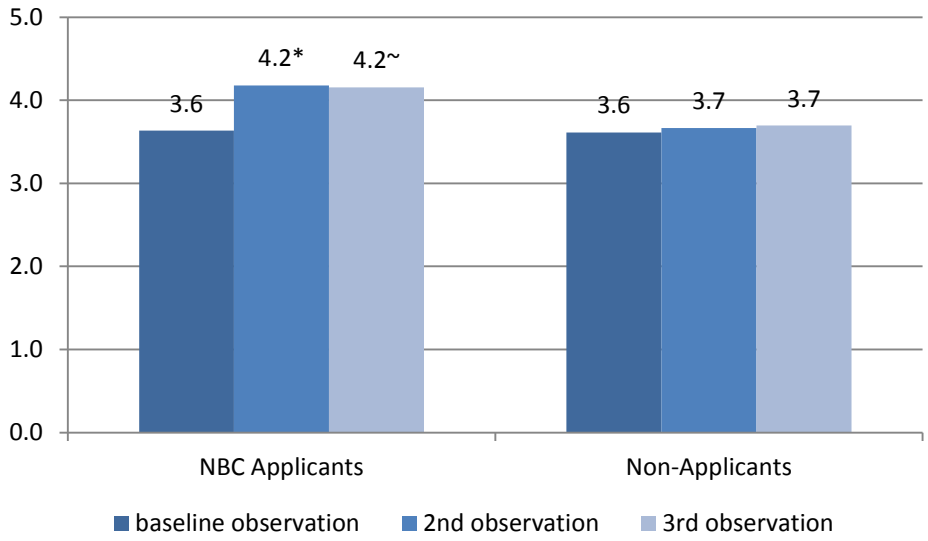
Table 8: Changes over time for the overall rating and subscale ratings for NBC applicants and non-applicants.

	NBC applicants			Non-applicants		
	Obsv. 1	Change: Obsv. 1 vs 2	Change: Obsv. 1 vs 3	Obsv. 1	Change: Obsv. 1 vs 2	Change: Obsv. 1 vs 3
Overall rating	3.77	0.20	-0.04	3.17	0.00	0.25
Lesson overview	4.17	-0.05	-0.05	3.40	0.30	0.15
Instructional overview	3.88	-0.01	0.06	3.28	0.22	0.11
Questioning	4.04	-0.19	-0.26	3.15	0.02	0.18
Classroom atmosphere	3.63	0.55*	0.53~	3.61	0.06	0.08
Higher-order skills	3.18	0.00	0.02	2.71	-0.25	0.21
Content knowledge	3.98	-0.21	0.05	3.33	0.07	0.00
Positive climate	4.43	0.01	-0.08	3.76	0.06	0.00
Implements instruction	3.99	-0.18	0.01	2.97	0.22	0.31
Assesses learning	3.63	-0.16	-0.10	2.86	0.42	0.33

NOTES: Scale ranges from 1 (low) to 5 (high). N=27 teachers. Significance was calculated using two-tailed t-tests of mean ratings for the baseline observation compared with each subsequent observation. *=difference is statistically significant using a 95 percent confidence level. ~=difference is statistically significant using a 90 percent confidence level.

Figure 5 shows changes over time in the average ratings on the *classroom atmosphere* subscale. The NBC applicants' average increased from baseline (3.6) to the second observation (4.2) and remained constant for the third observation (4.2). The improvement in the NBC applicants' average scores was statistically significant for the second observation relative to the baseline observation, as well as for the third observation relative to the baseline. The mean rating for the non-applicants remained similar at 3.6 to 3.7 for all three observations.

Figure 5: Average ratings on the "classroom atmosphere" subscale for NBC applicants and non-applicants, by timing of observation.



NOTES: scale ranges from 1(low) to 5(high). N=27 teachers. Significance was calculated using two-tailed t-tests of mean ratings for the baseline observation compared to each subsequent observation. *=difference is statistically significant using a 95 percent confidence level. ~=difference is statistically significant using a 90 percent confidence level.

To obtain the highest rating on the *classroom atmosphere* subscale, teachers must demonstrate the following:

- **Student involvement:** All of the students demonstrated interest and were engaged in the instructional activity.
- **Classroom management:** The classroom was well managed and totally orderly; there were no student disruptions which caused a loss of instructional time or impaired the learning environment.
- **Classroom culture:** The teacher has established a classroom culture in which all, or nearly all, of the students take initiative in discussions and activities; all students demonstrated respect for other students; all, or nearly all, demonstrated enthusiasm, confidence, persistence, and accuracy while solving problems.

In one observation of an NBC applicant with a rating of 5, the observer noted,

> All students were actively involved in every stage of the lesson for the full 100 minutes of class. They exhibited curiosity, confidence, persistence, responsibility, accuracy, and enthusiasm.

In another highly rated class, the observer described the classroom atmosphere by noting,

> No "down" time exists during this 60-minute class. All students are curious, persistent, confident, enthusiastic, and accurate in their work, and the environment is one of active thinking and learning from interaction among the content, the teacher, and the students. Students sit at science tables and discuss or share within a pair or threesome or across tables in larger groups.

This description seems to reflect what instruction would look like under the NBPTS "Architecture of Accomplished Teaching Helix," as shown in Figure 1 in the introduction. If the teacher is meeting the needs of each student at the place that student is, then all students should be engaged in the activities and behaving in an orderly manner. The classroom culture should also reflect student initiative, respect, and enthusiasm for learning.

Results: Changes in instructional quality for applicants with different baseline ratings

One potential limitation to examining changes in instructional quality for National Board applicants over time is the ceiling effect. Because NBC teachers begin with higher ratings for instructional quality than non-applicants at baseline, they may have limited room for improvement. We conducted additional analyses to examine this possibility, which are described in Appendix F.

We found no evidence that National Board applicants whose ratings at the baseline observation were in the bottom quartile demonstrate greater improvement over time than do applicants who whose baseline ratings are in the top quartile.

Results: Classroom context

Lastly, we examined differences in the *physical setting* subscale for NBC applicants and non-applicants. Even though these ratings of the

classroom context do not contribute to the ratings of instructional quality, they are important for understanding limitations in the types of activities that teachers may be able to conduct during their lessons.

The *physical setting* subscale is based on the following three items:

- **Classroom facilitates student learning:** This item considers the flexibility of student seating, the adequacy of utilities (e.g., electrical outlets), and whether flat top surfaces are available for conducting hands-on activities.
- **Classroom facility:** This is based on whether the classroom is adequate in size for the number of students, the adequacy of resources and equipment, and the availability of furnishings for activity-based instruction.
- **Classroom environment:** This item takes into account the availability of materials, textbooks and reference books, computers for student use, display of student work, and evidence of ongoing projects.

During our study, observers visited classrooms that ranged widely in classroom environment. In one observation that scored a 1 for *physical setting*, the observer noted,

> The room was sufficiently large to accommodate 25 students, but the furnishings included individual desks and no lab facilities. However, the most significant obstacle to high quality science was the requirement for teachers to move to another room for each class. Although there was a storage room adjacent to the classroom, the few pieces of science equipment observed were outdated, and in some cases, inoperable. Walls were devoid of anything related to science, and there were no displays of student work or projects.

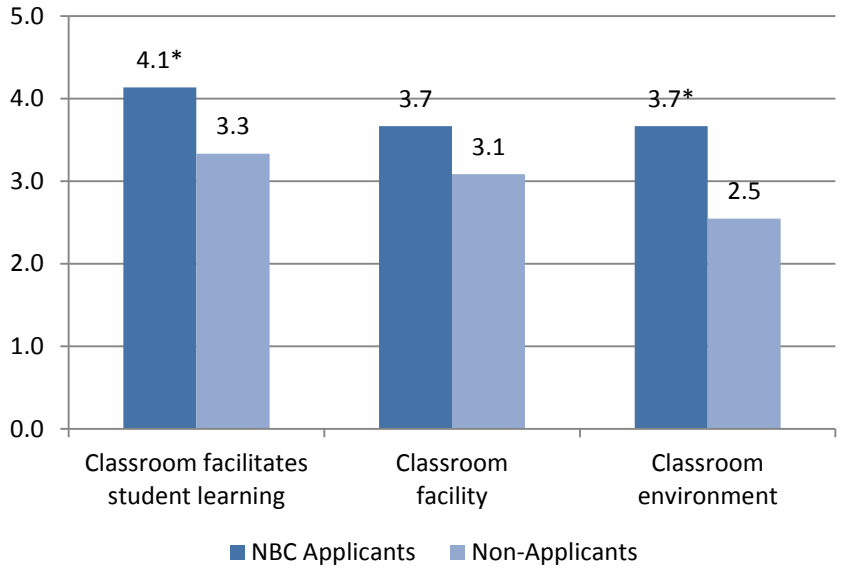
Such classrooms limit the types of activities the teacher could conduct. Conversely, during an observation that scored a 5, the observer noted,

> Students worked at tables using laptops, iPads, and TI-84 Plus calculators. Mathematics displays promoted learning.

As shown in Figure 6, there were differences by National Board applicant status for two of the three *physical setting* items. NBC applicants received higher ratings than non-applicants for "classroom

facilitates student learning" (mean of 4.1 versus 3.3, a difference of 0.8) and "classroom environment" (mean of 3.7 versus 2.5, a difference of 1.2). This suggests that National Board applicants taught in classrooms that were better designed for student learning and had access to more instructional resources than did non-applicants.

Figure 6: Average ratings on the three items of the "physical setting" subscale for NBC applicants and non-applicants.



NOTES: Scale ranges from 1 (low) to 5 (high). N=27 teachers. Ratings are from the baseline observation. Significance was calculated using two-tailed t-tests of mean ratings for NBC applicants compared with non-applicants. *=difference is statistically significant using a 95 percent confidence level. ~=difference is statistically significant using a 90 percent confidence level.

All of the control teachers except for two were from the same school as one of the National Board applicants. This means that the differences identified in the classroom context between applicants and non-applicants are occurring within the same school. These findings may suggest that National Board applicants are more resourceful in organizing their classrooms or obtaining the necessary resources to support productive learning than are non-applicants.

This page intentionally left blank.

Student outcomes

As described in the previous section, we used qualitative data from classroom observations to address our first research question:

1. Does the National Board certification process influence teachers' classroom practices?

The goal of the statistical analysis of student test scores described in this section is to answer our remaining questions:

2. Are National Board–certified teachers more effective than other teachers?

3. Are applicants who attain National Board certification more effective than applicants who do not?

4. What effect, if any, does the National Board certification *process* have on teacher effectiveness?

To answer the different questions, we compare different groups of teachers. We explore the first question, which asks whether National Board certification is a good signal of teacher effectiveness, by comparing the effectiveness of National Board–certified teachers with teachers who are not certified. The second question, which considers the effectiveness of National Board certification as a screening process, is answered by comparing teachers who apply for and achieve certification with those who apply for but do not achieve it. The third question addresses the professional developmental properties of the National Board certification process itself, by comparing the effectiveness of individual teachers against themselves at different stages (before, during, and after) in their application process.

In each case, we will examine the evidence of teacher effectiveness as measured by student posttest scores on the ACT and the PLAN standardized tests.

Methods: Estimation model

We will use an "education production function" approach to relate school, teacher, and student-level characteristics to the outcome, with the base statistical model being a standard linear regression model. Each observation represents an individual student linked to his or her current subject-area teacher (or set of subject-area teachers, in the case of students who had multiple teachers between pretest and posttest; see the "Description of the Data" chapter). All models correct the standard errors for clustering of the data by teacher.

Outcome (dependent) variables

For each of these three research questions, the outcome variable is a student's test score. One set of models, which we refer to as the "PLAN to ACT" analysis, use the student's ACT score as the outcome, with the student's previous PLAN score as the prior test score. A second set of models use the student's PLAN score as the outcome, with the student's previous EXPLORE score as the prior test score. We refer to this second model as the "EXPLORE to PLAN" analysis. Separate models are run for each subject: math, English, and science. We also run a combined model that includes all of the subjects, with additional variables to indicate whether the observation outcome represents a math, English, or science test score.[8] Results are also presented separately for Kentucky and CPS.

One difference between our study and other studies in this literature is that we do not have an annual student achievement measure. In Kentucky, students typically take the EXPLORE at the beginning of 8th grade, the PLAN at the beginning of 10th grade, and the ACT at the end of 11th grade. Thus, depending on the analysis, the prior test score occurs three to four semesters before the posttest outcome. Because there are multiple semesters between the prior score and the outcome, and these are high school students who may switch teachers

[8] So that we can compare scores across subject areas, we standardize all test scores used in our models by subtracting the national-level subject-specific mean from the student's score and dividing by the national-level subject-specific standard deviation.

from semester to semester, each student-level observation will involve more than one teacher.

In Kentucky, we observe the student's course-taking each semester; so, for a given subject, there will typically be three or four teachers between the pretest and the outcome. In Chicago, the test-taking schedule is different, in that students typically take the EXPLORE in 9th grade rather than 8th. Additionally, in Chicago, core courses typically run for a full year; because we only observe student course-taking on a year-by-year basis, rather than each semester, there will be at most two teachers per student, per subject, between the pretest and the outcome for CPS analyses.

Explanatory variables

A challenge in estimating teacher effectiveness using longitudinal data systems, as we do here, is that neither teachers nor students are randomly assigned to their classrooms, or to their schools. Education-minded parents choose housing taking school quality into account; teachers choose where to work based in part on the school's quality; the most effective school leaders find ways to recruit early to obtain the best candidates; and once in their schools, principals assign students to teachers thoughtfully, not at random.

As a result, there likely are systematic differences in student and teaching assignments that affect test scores, but that have nothing to do with National Board certification. Because of this challenge, for each analysis we use a variety of statistical controls and estimate five different regression models to get a fuller picture of the likely true effect of National Board certification on student test scores.

Model 1 is our baseline model. It includes the student's prior score (the EXPLORE score in the case of models with the PLAN as the outcome, and the PLAN score in the case of models with the ACT as the outcome), by subject, to control for past student achievement. It also includes student age, the number of student absences (KY only), and standard demographic indicators for racial/ethnic background, gender, FRL eligibility, special education status (IEPs), and ESL status (KY only). Controlling for these observable student characteristics helps level the playing field when we compare student outcomes and attribute differences to teaching effectiveness. Model 1 also includes a control for the number of years of experience for each teacher for

CPS; for Kentucky, it includes a proxy for experience, given by the number of years the teacher appears in the dataset.[9]

Model 1 likely overstates the true NBC effect, because it does not take into account all of the differences in students that may be present between NBCTs and the comparison teachers. In addition, the model does not account for differences across schools in the contributions schools make to student performance, including, the contributions of school leaders and administrators, instructional materials, and other students. But it does provide us with a best-case, baseline estimate of teacher effectiveness, comparing NBCTs to other teachers across the district or state, after controlling for the characteristics of students assigned to each teacher.

Model 2 adds to model 1 a set of school characteristics, to control for across-school differences. Our school-level variables include total enrollment, student-teacher ratio, racial/ethnic composition of the student body, and percentage of the student body FRL eligible. We also include, at the district level for Kentucky, the student-administrator ratio and per-pupil spending, as well as the urban-centric locale code (urban, suburban, town, or rural).[10] We also include the school-level average pretest score (the EXPLORE for the analysis with the PLAN outcome, and the PLAN for the analysis with the ACT outcome) in English, math, and science, as a measure of the school's overall achievement level. In model 2, our comparison is between NBCTs and other teachers in similar schools, controlling for characteristics of each student assigned to them.

---

[9] For Kentucky, if the student was assigned to multiple teachers, or the teacher was unknown, we treated the teacher experience proxy variable as missing data, and flagged the observation. For the average incoming prior test score, we calculated separate averages for students assigned to "BLOCK," "MISSING," and "MULTIPLE," respectively. In Chicago, it is the overall average regardless of why the student does not have an individually identified teacher. For CPS, we also include "experience squared." This variable accommodates the nonlinear relationship between experience and teacher effectiveness.

10. These variables are not included in the CPS model since it is a single district.

Model 3 takes a step back and adds to model 1 the average prior test score for the group of students assigned to each teacher. Including this variable better accounts for within-school differences in how students are assigned to teachers that may be correlated with student outcomes. While model 1 controls only for the characteristics of individual students, model 3 takes into account the overall prior performance of students, which can affect instructional challenges in the classroom.

Model 4 adds to model 1 both the school characteristics used in model 2 and the average prior test scores of students assigned to each teacher, used in model 3. This provides us with an estimate of the NBC effect to address the nonrandom assignment of students both across and within schools.

Our final model, model 5, replaces the set of student characteristics in model 4 with a set of school-level fixed effects. The school fixed-effects model provides a stronger control for differences across schools that may affect our measurement of teacher effectiveness, because it provides a way to account for time- and subject-invariant school-specific factors that influence student performance that we otherwise cannot observe in our data.

In general, we expect model 5 to provide our most conservative estimate of the effectiveness of NBCTs compared with other teachers. However, this model may actually understate the difference in effectiveness between NBCTs and other teachers, because teachers also sort themselves across schools. Indeed, unlike model 1, which provides an (likely overstated) estimate of the effectiveness premium of NBCTs compared with a typical teacher in the system, model 5 provides an estimate of the National Board effect in comparison to a typical teacher in the same school. Because teachers within a single school are generally more similar to one another than to other teachers from across the district or state, all else equal, this teacher self-sorting likely will reduce estimates of the relative effectiveness of National Board–certified teachers.

National Board status indicators

After controlling for the variables described above, the covariates of interest will be the set of indicators that summarize a teacher's status with respect to the National Board certification process. The precise

set of indicators will differ depending on the research question being addressed.

Methods: Signaling effect

To test for a signaling effect of National Board certification, we compare the test scores of students who had one or more National Board–certified teachers between the pre- and the posttest with scores of students who had no NBCTs between the tests. If National Board certification is an effective signal of teaching quality, then students taught by certified teachers should perform better on tests than students taught by non-certified teachers.

We will estimate a model that includes an indicator variable that equals 1 if the student had a National Board–certified teacher in the tested subject area in any semester (or any year, for CPS students) in which we observe the teacher, and 0 otherwise. This model provides a comparison of the performance of students who had at least one National Board–certified teacher between the pre- and the posttest with the performance of those students who had none.

Methods: Screening effect

To test for a screening effect, we compare the performance of students who had teachers who will ever achieve certification ("ever certified"), whether before, during, or after the timeframe of our analysis, with the performance of students who had teachers who have or will apply but not achieve certification ("never certified"). If the National Board certification process is an effective screening device for high-quality teachers, then students taught by "ever certified" teachers should perform better on tests than students taught by "never certified" teachers.

For the screening model, the teacher status variable will indicate the total number of semesters (or years, in the case of CPS students) that the student had a National Board–certified teacher. We will include three variables for both the EXPLORE to PLAN and the PLAN to ACT analyses: the number of semesters taught by an "ever certified" teacher, the number of semesters taught by a "never certified" teacher, and the number of semesters taught by a "never certified-withdrawn" teacher. This formulation allows us to distinguish the

program effects by the amount of instructional contact the student had with a particular type of teacher.

To estimate the screening effect size, we can ask the following question: what would be the effect on a student's test score if we replaced a "never certified" teacher with an "ever certified" teacher? The quantity we are looking for will be the difference between the coefficient on the status indicator for "ever-certified" teachers and the coefficient on the status indicator for "never certified" teachers.

Methods: Human capital effect

To estimate the effect of the certification process itself on teacher effectiveness, we want to compare the student performance of teachers who have completed the application process ("past applicant") with the student performance of these same teachers when they were applicants ("current applicant"), and with the performance of their students before they started the certification process ("future applicant"). If the National Board certification process itself is effective professional development, then we should expect to see a positive coefficient on the "past applicant" indicator—implying that students of past applicants have higher levels of achievement than students of future applicants do.

Additionally, some previous studies have found evidence that current applicants may be less effective than either past or future applicants. We can use this model to investigate any such potential effects in our sample.[11]

11. Note that for human capital models, for both Kentucky and Chicago, we define the application status variables as spanning one academic year (rather than one semester, as is the case with Kentucky signaling and screening models). The models therefore include one teacher per student per year in a subject. For Kentucky students who have more than one teacher in a school year, we created a special "MULTIPLE" teachers category for variables that depend on the identity of the teacher. We adopt this approach because of identification concerns with respect to the teacher fixed effects model that we use for the human capital effect estimates.

Teacher fixed effects in human capital models

In human capital models, we include a set of teacher fixed effects for the current teacher in a subject.[12] The idea behind this approach is that our basic model includes teacher fixed effects and the NBC status variables that for an individual teacher will change over time as the teacher moves through the application process. Therefore, we identify only the effect of going through the process for teachers who are changing status during the timeframe in which we observe student data. We are estimating the human capital effect by comparing the same teacher with himself or herself over time.

Results: Signaling effect

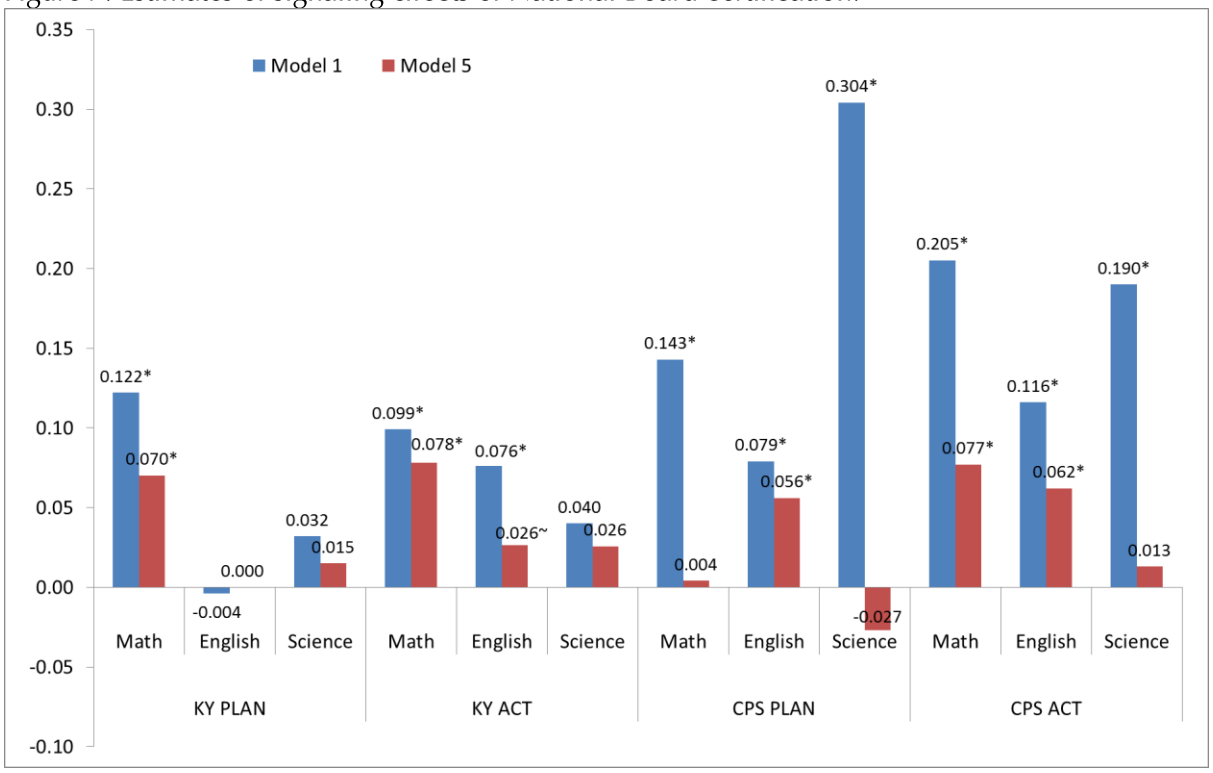
This section and the next two sections present the results of the statistical analyses of student test scores to estimate any signaling, screening, and human capital effects of the National Board certification process. For each analysis, we estimated a series of statistical models incorporating a range of different covariates. Results presented for model 1, with the fewest controls, include a set of student characteristics and teacher experience (a proxy variable—the number of years the teacher appears in the dataset—in the case of Kentucky). Results presented for model 5 include controls for student characteristics, teacher experience, teacher incoming students' average prior test score, and a school-level fixed effect.

We compare the results from model 1 with model 5 in the body of the report to provide an estimate of the range of effect sizes, depending on the statistical controls included in the model. Complete results for all models (1 through 5) are presented in Appendix E.

12. For EXPLORE to PLAN analyses, the current teacher is the grade 9 teacher because the PLAN tests are administered in the fall of grade 10. For PLAN to ACT, the current teacher is the grade 11 teacher because the ACT is administered in the spring of grade 11. To reiterate, for Kentucky students who have more than one teacher in a school year, we created a special "MULTIPLE" teachers category for that year. In both Kentucky and CPS, we include fixed effects for both the 10th and 11th grade teachers in the PLAN to ACT analysis.

To estimate the signaling effect, we compared teachers who currently are National Board certified with those who are not. Figure 7 summarizes our estimates of the signaling effect, by subject. To measure the effect size of having an NBCT, the indicator equals 1 if the student had a National Board–certified teacher in any semester or school year in the tested subject area between the pre- and posttest, and equals 0 if the student did not. The coefficient can be interpreted as the effect size on the outcome variable (i.e., the number of standard deviations of change in the outcome variable) associated with having at least one National Board–certified teacher in that subject between the pre- and posttest.

Figure 7: Estimates of signaling effects of National Board certification.



NOTES: N=80,253 for Kentucky PLAN, N=114,004 for Kentucky ACT, N=69,741 for CPS PLAN, and N=48,546 for CPS ACT. Significance was calculated using multiple regression models for the effect of having an NBCT in any semester or year on student test scores. *=difference is statistically significant at the .05 level. ~=difference is statistically significant at the .1 level. See Appendix E, Table 13, Table 14, and Table 15 for details of the regressions.

For Kentucky math students, there is a positive and statistically significant, although small, effect on both ACT and PLAN scores of having at least one NBCT in the subject area between the pre- and posttest.

The effect size ranges from a .070 to .122 standard deviation increase in both ACT and PLAN math scores.[13] For English on the ACT outcome only, there is a positive effect of 0.076 in model 1. However, the signaling effect is not statistically significant at conventional levels in model 5 for English.

For CPS, results in model 1 are positive and statistically significant for all subject areas in both the PLAN and the ACT analysis, with effect sizes ranging from .079 in the English PLAN analysis to .304 in the science PLAN analysis. When additional control variables are added in model 5, statistically significant effects are present only for English on the PLAN outcome (effect size of .056), and for math and English on the ACT outcome (effect sizes of .077 and .062, respectively).

Results: Screening effect

To estimate the screening effect, we compare student test scores of teachers who currently hold or in the future will hold National Board certification with test scores of teachers who have applied for certification in the past, or will do so in the future, but who do not achieve certification. As mentioned, we measure the screening effect by the difference between the coefficient on the status indicator for number of semesters/years with an "ever certified" teacher and the coefficient on the status indicator for number of semesters (or years) with a "never certified" teacher.

In Figure 8, the effect size should be interpreted as the change in score that would be brought about by replacing one "never certified" teacher with one certified teacher.[14] The results for Kentucky indi-

13. An effect size of .07 implies that if Kentucky reassigned the median student to an NBCT, the student would move from the 50.0th to the 53.7th percentile on the ACT math.

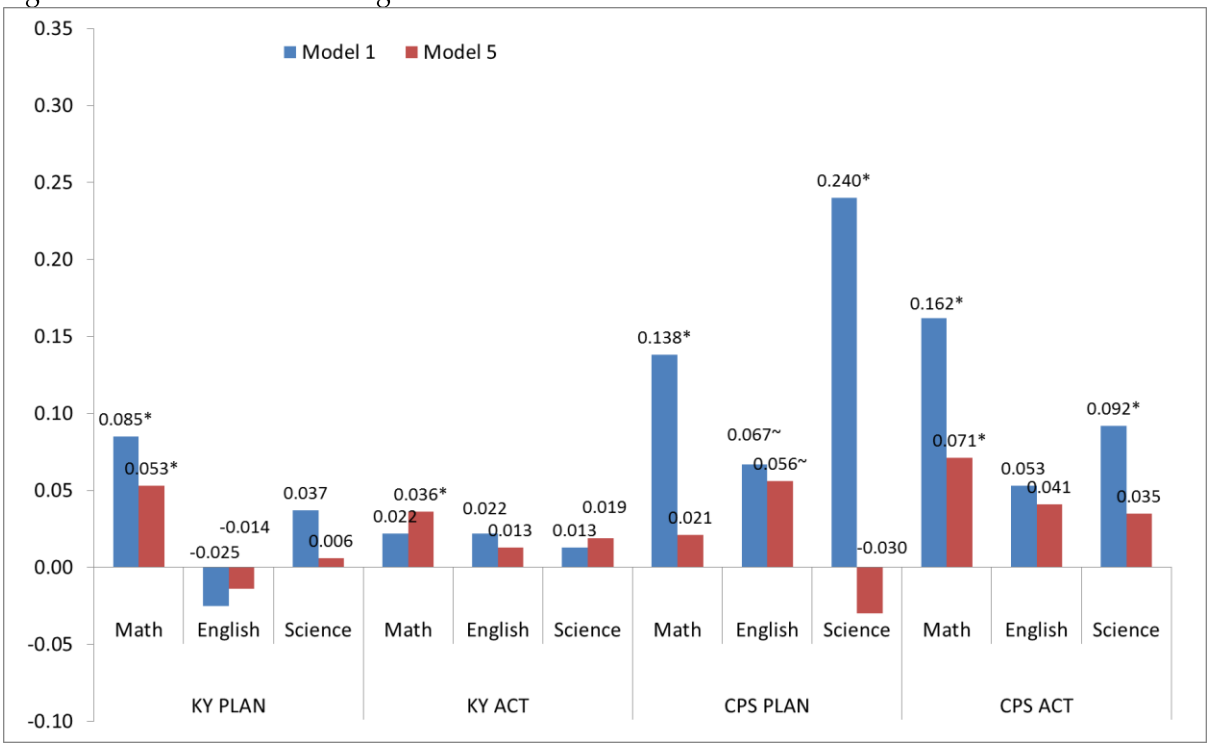
14 Appendix E, Table 17 (math), Table 18 (English), and Table 19 (science) present additional results for various specifications of the screening model, by subject, outcome, and National Board status variable. To interpret the coefficients in these tables, note that the comparison group (the omitted group) consists primarily of nonparticipating teachers, plus a few participants whose ultimate status we do not observe.

cate a small but statistically significant screening effect for math on the PLAN and the ACT outcomes, with effect sizes ranging from .036 to .085. This suggests that, in Kentucky, the National Board certification process does screen in math teachers who are slightly more-effective compared with those who do not achieve certification.

In Chicago, the differences between successful and unsuccessful current, future, or past applicants are a mix of nonsignificant and statistically significant positive effects of the NBC process. In model 1, the results are positive and significant in all subject areas for both the PLAN and the ACT outcomes (except for English on the ACT outcome), with effect sizes ranging from .067 (English on the PLAN outcome) to .240 (science on the PLAN outcome). In model 5 there are positive effects in English (with effect sizes of .056 on the PLAN outcome and .041 on the ACT outcome), and in math on the ACT outcome (effect size of .071).[15]

15. As a sensitivity test, we also estimate the screening model by comparing applicants who achieved with applicants who did not achieve before they ever entered the application process (i.e., when they are pre-applicants). This tests for the presence of a screening effect before teachers' practices may be influenced by the certification process. In Kentucky, the results are similar for almost all models, except that the effect for math on the PLAN outcome is no longer statistically significant. In CPS, the effects for English are no longer statistically significant for the PLAN and ACT outcomes. In addition, the effect for math is no longer statistically significant on the ACT outcome.

Figure 8: Estimates of screening effects of National Board certification.



NOTES: N=80,263 for Kentucky PLAN, N=114,019 for Kentucky ACT, N=69,741 for CPS PLAN, and N=48,546 for CPS ACT. Significance was calculated using multiple regression models for the effect of having an NBCT in any semester or year on student PLAN and ACT scores. *=difference is statistically significant at the .05 level. ~=difference is statistically significant at the .1 level. See Appendix E, Table 17, Table 18, and Table 19 for details of the regressions.

Results: Human capital effect

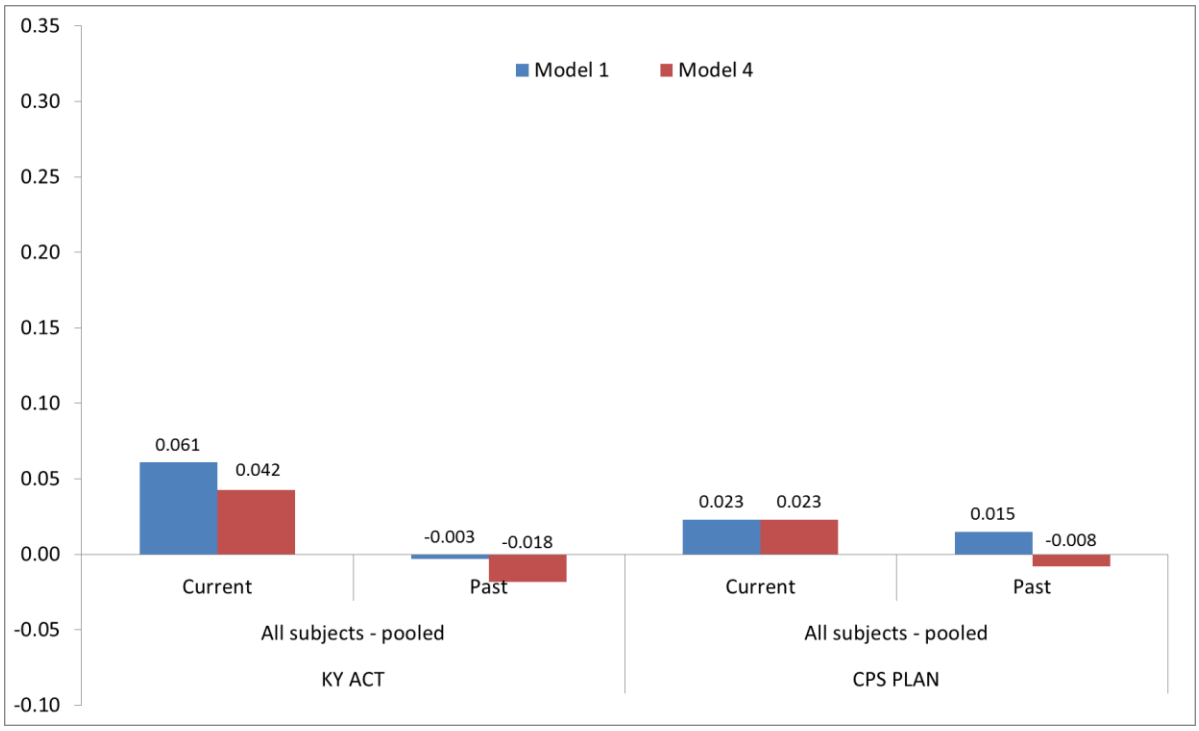
To estimate the human capital effect, we compare the same teacher with himself or herself over time as the teacher moves from future applicant to current applicant to past applicant. The model includes National Board status indicators for whether the teacher is currently in, or has in the past participated in, the National Board application process, along with a current teacher fixed effect, a school-level fixed effect, and student characteristics.

The omitted category is "future applicant," so the coefficient ("effect size") should be interpreted as the change in outcome score (in standard deviations from the national mean) resulting from having a teacher who is a current (or past) NBC applicant relative to having the same teacher at a stage in her or his career when she or he had not yet applied for certification. The coefficients should therefore

pick up any effect on test scores from teachers who have gone through (past applicant), or are going through (current applicant), the National Board certification process. The results of all subject areas are pooled due to the small number of teachers who change status in the certification process during the timeframe of the analysis.

Figure 9 summarizes the results. We find little evidence of a human capital effect; students of past or current applicants do not, in general, perform differently from students of the same teachers before they had applied for National Board certification (future applicants). The effect sizes on both the current and past applicant indicator variables are small and not statistically significant in Kentucky or Chicago for the PLAN and the ACT outcomes.

Figure 9. Estimates of human capital effects of National Board certification.



NOTES: N=80,163 for Kentucky PLAN, N=113,923 for Kentucky ACT, N=69,741 for CPS PLAN, and N=48,546 for CPS ACT. Significance was calculated using multivariate regression models for the effect of having a teacher in the NBC application process on student PLAN or ACT scores. *=difference is statistically significant at the .05 level. ~=difference is statistically significant at the .1 level. See Appendix E, Table 21 for details of the regressions.

This page intentionally left blank.

Conclusions

Key findings

In summarizing and evaluating an accumulation of research, the National Academy of Sciences concludes, "The evidence is clear National Board certification distinguishes more effective teachers from less effective teachers with respect to student achievement" (National Research Council, 2008, p. 179). But the extant literature leaves understudied, and unresolved, whether undergoing the certification process itself also improves a teacher's effectiveness.

The results of this study go beyond estimation of signaling and screening effects by examining the evidence of improvement in teacher effectiveness as teachers progress through the NBC process and ultimately become a National Board–certified professional. We also examine differences in performance growth between successful and unsuccessful applicants to determine whether the certification process is an effective screening tool.

Moreover, the signaling literature, which compares outcomes of students with and without NBCTs, focuses almost exclusively on statistical comparisons in just two states, Florida and North Carolina. Another contribution of the current study is that it uses data from two new locales, the state of Kentucky and the city of Chicago, that together include rural, suburban, and large urban districts. Both locales are strong supporters of the NBC process, as evidenced by the proportions of teachers who hold certification in those locations. The statistical analysis focuses on student outcomes in English, science, and math at the high school level. To complement the large-scale longitudinal analyses of student achievement, we conducted classroom observations to examine changes in the quality of instruction over time, comparing applicant teachers as they progressed through the NBC process and non-applicants.

We find that when NBC applicants are observed as they are beginning the certification process, they already have higher ratings of instructional quality than non-applicants. These results are seen in

teachers' overall ratings as well as ratings on six of nine or our rubric subscales of teaching quality: *lesson overview, questioning, content knowledge, positive climate, implements instruction,* and *assesses learning.*

However, there are few such changes in instructional quality. There are no significant differences from the first observation to the second or third observation in the overall rating or the ratings for eight of the nine subscales for both NBC applicants and non-applicants. The one area where we did find an improvement over time for NBC applicants is the subscale for *classroom atmosphere*. This subscale takes into account student involvement, classroom management, and classroom culture, reflecting many of the ideals represented in NBPTS's "Architecture of Accomplished Teaching Helix."

In the statistical analysis of student outcomes, we find small signaling effects of NBC in several subject areas and outcomes. Students who have at least one National Board–certified teacher between taking the pretest and the posttest tend to score slightly better on the posttest compared with students who do not have a certified teacher. We also find small screening effects of the NBC process. The certification process does seem to identify slightly more effective teachers compared with those who do not achieve certification. We find little evidence, however, of a human capital effect of undergoing the NBC process for Kentucky or Chicago teachers. Students of past or current applicants do not, in general, perform differently on the posttest than do students of the same teachers before they had applied for National Board certification.

Limitations

There are several limitations of the study that should be taken into consideration when reflecting on its findings. First, there may be a ceiling effect on the growth of National Board applicants. National Board teachers start out with higher ratings in instructional quality and higher levels of student performance at the beginning of the certification process, and so have less room for improvement than do non-applicants.

Second, ours is a relatively small sample of National Board applicants in both the classroom observations and the statistical analysis. This

makes it difficult to detect statistically significant changes over time. In the human capital model, for example, we had hoped to measure possibly differential changes over time for teachers who achieved, did not achieve, and withdrew from NBC during the study period. However, there were an insufficient number of each.

Third, there are limitations with the timing of the classroom observations. For NBC participant teachers in the study, it is not possible to observe them before they have had any involvement in the National Board certification process, because we cannot know teachers' intentions until they apply. While we tried to observe teachers as close to their becoming applicants as possible, the baseline observations may still reflect some exposure to the certification process.

Some teachers enroll in NBPTS's preparatory Take One! activity before they apply for certification, for example, so they may already be making changes to their teaching practices when they are new applicants. In addition, the last of the classroom observations was conducted three semesters into the NBC process. At this point, applicants typically have completed most certification activities (e.g., submitting portfolio entries), but they may not be entirely finished. Thus, applicants may continue to change their teaching practices in response to what they are learning from certification beyond that last observation. This is especially true for applicants who may have been unsuccessful on their first attempt at certification and go on to reapply. In addition, both NBC applicants and non-applicants may have been exposed to other types of professional development related to instructional strategies (particularly in regards to formative and summative assessment), which could influence instructional quality during the observation period.

Last, there are several limitations to the statistical analysis. The data collected for analysis included a limited number of characteristics for students, teachers, and schools. Our description of the data indicates that differences exist between students and schools with NBCTs and those without. This suggests there is selection bias in how teachers are distributed among and within schools. Our statistical models control for some of these differences, but there may be other unobserved factors contributing to changes in student outcomes that are also corrected for with a teacher's NBC status.

Another analytic limitation is the time lag between the pretest and the posttest makes it difficult to attribute changes in student achievement to an individual teacher. Our statistical models take into account the effect of all the teachers for that specific subject that students had between the two tests, but this may diminish the impact of any one teacher. If a student had a high-quality teacher in the pretest year and a low-quality teacher in the posttest year, we might expect the student's growth to be lower than if the student had the high-quality teacher for the full duration.

Implications for future research

The results of our study have several important implications for future research on National Board certification. Our analysis of classroom observations suggests that it is important for studies to look beyond traditional outcomes of student achievement (test scores), and also consider teacher outcomes such as quality of instruction. While we found that National Board applicants demonstrate improvement in classroom atmosphere over time, we need more context for understanding what caused these changes in classroom atmosphere. Surveys or interviews could be conducted with National Board applicants to better understand how and why they may have changed their classroom practices as a result of participating in the certification process. Future research could also examine portfolios of students' work for teachers before and after participating in the certification process. This would provide evidence of whether NBCTs are more effective at challenging students and at creating individualized assignments based on where students are at—skills that are emphasized in NBC.

The statistical analysis of student outcomes could be extended by obtaining a larger sample, by adding either more years of data or additional locales. This would allow for a more nuanced examination at various stages of the certification process. For example, do outcomes differ for applicants who achieve certification upon their first attempt compared with applicants who achieve it after two or more application cycles? With a larger sample, it would also be possible to examine differences by certification type (e.g., math teachers with Generalist certifications compared with teachers with Mathematics/Adolescence and Young Adulthood certifications).

Implications for practice

Large investments have been made in the development of the National Board for Professional Teaching Standards certification program. As of September 2005, the National Science Foundation and the U.S. Department of Education had appropriated more than $149 million dollars to it, and nongovernment funders had spent an additional $261 million (Cohen & Rice, 2005). Additionally, there are ongoing costs incurred by applicants or (more typically) their sponsoring school systems.

As a result of these investments, there is a great deal of interest in identifying and measuring the full value to education systems of encouraging teachers to obtain National Board certification. The "signaling" value of certification has been demonstrated, and the long-term benefits to improvements in the workforce have been postulated, but there also is interest in measurement of more immediate effects of certification on the instructional effectiveness of participants in the program.

Although its findings are modest, this study contributes to better understanding the full benefits of encouraging National Board certification, which may inform future budget decisions by districts or state departments of education about subsidization of the NBC process. Although the cost of the NBC program has been considerable, in fact it is much less expensive than raising teacher salaries enough across the board to make up for years of salary declines (in real terms and relative to other professions requiring similar skills) that may have weakened the quality of new entrants to the profession and the teaching workforce generally (e.g., Burke et al., 2004).

Given that the National Board certification process has repeatedly demonstrated the ability to distinguish between more- and less-effective teachers, school systems should think about how to make good use of this tool. For example, school systems could use National Board certification as a gatekeeper for tenure, implemented at a later point in the teaching career path than the criteria most school systems currently use for those decisions. School systems could also link certification to compensation. Over time, pay differences would be expected to encourage certified teachers to stay in teaching, and un-

successful applicants to leave, creating openings for new, more promising entrants.

Appendix A. Leadership by Design

Science classroom observation instrument

SCIENCE CLASSROOM OBSERVATION INSTRUMENT Code #:
Leadership by Design (NBC Version)

Level/Class ____ Lesson Title _____ Total # Students _____ Gender #: M____F ____
Minority____ # Inclusion____ Length of Observation _____

Learning Objective of the Lesson_____

I. LESSON OVERVIEW

A. Learning Objective of the Lesson (Mark all that apply)
☐ Clearly communicated by the teacher using multiple means ☐ Communicated orally only ☐ Communicated in writing only ☐ Student activities consistent with the lesson objective(s) ☐ Student activities not consistent with the lesson objective(s) ☐ Lesson objective communicated but not clear ☐ Lesson objective not communicated

B. Major Instructional Resource used in the lesson observed - Mark with a "1" the primary/predominant resource influencing instruction and following numbers (2, 3, etc.) if more than one observed - Sections B, D, E.
☐ Textbook ☐ Other Print Materials (worksheet, manual, etc.) ☐ Technology Based Presentation Media (Power Point, Smart Board, etc.) ☐ VCR, Overhead Projector ☐ Hands-on/manipulative materials (laboratory materials) ☐ Calculators ☐ Computer ☐ Graphing Calculator ☐ Technology Probes ☐ Centers ☐ None of Above

C. Content Delivery (Mark all that apply)	D. Place in Instructional Sequence	E. Seating Arrangement for Lesson
☐ Age/grade level appropriate	☐ Introduction of new concept	☐ Whole group
☐ Content presented is accurate	☐ Develop conceptual understanding	☐ Small groups working on same task
☐ One or more content errors	☐ Apply concept to new situation	☐ Small groups working on different task
☐ Student misconception not corrected	☐ Review concept or procedure	☐ Individuals working on same task
	☐ Assess student understanding	☐ Individuals working on different tasks

II. INSTRUCTIONAL OVERVIEW - Mark with a "1" the primary/predominant resource influencing instruction and following numbers (2, 3, etc.)if more than one observed – Sections A & B.

A. Instructional Strategy
☐ Teacher lecture ☐ Teacher demonstration ☐ Teacher-led discussion ☐ Individual assistance
☐ Student presentation ☐ Small group discussion ☐ Student Investigation ☐ Student Experiment
☐ Using a Model to Teach a Science Concept

B. Student Activity
☐ Listening to/observing teacher presentation ☐ Participating in discussion (teacher led or small group)
☐ Conducting investigation ☐ Conducting student or teacher initiated experiment ☐ Print-based Activity: Reading, answering questions ☐ Working on written assignment (science notebook, writing a lab report, etc.) ☐ Taking a test
☐ Using education software program ☐ Using technology for research ☐ Using computer for Inputting/analyzing data
☐ Student Presentation and/or Listening to Other Student Presentation ☐ Developing/Using a model to learn or clarify a concept

III. QUESTIONING

A. Quality of the Questions (Mark only one)
☐ Questions were mostly convergent focusing on factual recall
☐ Questions were mostly divergent and stimulated broad student responses
☐ Appropriate balance of divergent and convergent questions
☐ No questions were asked by teacher or posed through the activity being conducted

B. Questioning Techniques (Check all that apply)
☐ Students are encouraged to ask questions of each other and/or teacher; ☐ Questions stimulated higher level and divergent thinking ☐ Appropriate wait time ☐ All students have an opportunity to respond ☐ Most students have an opportunity to respond ☐ Only a few students have an opportunity to respond ☐ Teacher provides appropriate feedback to student responses

IV. CLASSROOM ATMOSPHERE (Mark one response in each section A and B)

A. Student Involvement (Check only one)

- ☐ All or nearly all students demonstrate interest and were engaged
- ☐ Majority of students demonstrate interest, were engaged
- ☐ Approximately equal numbers of students interested/engaged and not interested/not engaged
- ☐ Majority of students uninterested or apathetic; generally not engaged
- ☐ Nearly all of the students were uninterested and not engaged

B. Classroom Management (Check only one)

- ☐ Classroom orderly, no student disruptions which impaired learning environment
- ☐ Classroom generally orderly but some student disruptions which required corrective action
- ☐ Classroom disorderly, frequent student disruptions which seriously impaired the learning environment

C. Classroom Culture (Check all that apply)

- ☐ Curiosity ☐ Cooperation with teacher and/or other students ☐ Persistence ☐ Responsibility
- ☐ Confidence in ability to "do" science ☐ Enthusiasm for learning ☐ Objectivity in analyzing data ☐ Accuracy
- ☐ Use of critical thinking skills

V. ANALYSIS OF INSTRUCTION LEADING TO THE DEVELOPMENT OF HIGHER ORDER SKILLS

A. Amount of Student Investigation/Research (Mark only one box for section A)

- ☐ Students are engaged in an investigation/research which may include skills 1-7, however the emphasis is on higher level skills 8-16.
- ☐ Students are engaged in investigation/research in which the focus of lesson is on the basic process skills 1-7.
- ☐ Students are not involved in any type of investigation/research involving hands-on or laboratory activity.

A. Level of Student Investigation/Research (Mark only one box for Section B)

- ☐ Students design and carry out an experiment to solve a problem initiated by a teacher or student question.
- ☐ Students are investigating a science phenomenon using a preplanned activity which requires the collection and analysis of data to solve a problem or create a product.
- ☐ Students are investigating a science phenomenon using a preplanned activity which provides a definitive procedure and requires a specific response to be correct. Does not necessarily involve collection and analysis of data.
- ☐ Students are not involved in any type of investigation/research involving hands-on or laboratory activity

C. Scientific Skills Being Developed (Check all skills which are introduced and/or developed in the observed lesson)

Basic Skills (Mark all that are observed)

- ☐ 1. Observing ☐ 2. Measuring ☐ 3. Classifying ☐ 4. Inferring ☐ 5. Predicting ☐ 6. Communicating
- ☐ 7. Investigating (Basic Level)

Higher Level Skills (Mark all that are observed)

- ☐ 8. Investigating (Involves Analysis of Data) ☐ 9. Designing Experiments ☐ 10. Formulating Hypotheses
- ☐ 11. Conducting Experiment ☐ 12. Collecting Data ☐ 13. Interpreting Data
- ☐ 14. Forming Conclusions ☐ 15. Evaluating Data ☐ 16. Interpretive Discussion

VI: TEACHER DEMONSTRATES APPLIED CONTENT KNOWLEDGE (Mark one response for each section)

A. Communication

- ☐ Consistently used accurate and effective communication; vocabulary is clear, correct and appropriate.
- ☐ Generally used accurate and effective communication; occasional use of inappropriate vocabulary.
- ☐ Consistently used inaccurate and ineffective communication and/or inappropriate vocabulary.

B. Connects Content to Life Experiences (Mark one response in this section)

- ☐ Consistently connected most content/procedures/activities with relevant life experiences.
- ☐ Connected some content/procedures/activities with relevant life experiences.
- ☐ Rarely or never connected content/procedures/activities with relevant life experiences.

C. Instructional Strategies Appropriate for Content and Contribute to Student Learning

- ☐ Used instructional strategies that were clearly appropriate for the content/processes of the lesson.
- ☐ Used instructional strategies that were generally appropriate for the content/processes of the lesson.
- ☐ Used instructional strategies that were questionable or inappropriate for the content/processes of the lesson.

D. Guides Students to Understand Lesson Content from Various Perspectives to Extend Understanding

- ☐ Provided multiple opportunities for students to consider content from a different context or perspective.
- ☐ Provided a single opportunity for students to consider content from a different context or perspective.
- ☐ Never provided an opportunity for students to consider content from different context or perspective.

VII: TEACHER CREATES AND MAINTAINS LEARNING CLIMATE (Mark one response for each section)

A. Communicates High Expectations

- ☐ Significant/challenging lesson objectives; teacher consistently communicates confidence in students' ability to achieve.
- ☐ Challenging objectives; some communication of confidence in students' ability to achieve.
- ☐ Minimal objectives for students; rarely or never communicates confidence in students' ability to achieve.

B. Establishes a Positive Learning Environment

- ☐ Clear conduct standards; awareness of student behavior; responded appropriately/respectfully.
- ☐ Conduct standards but some inconsistency in monitoring and response to student behavior.
- ☐ No established conduct expectations; minimal or no monitoring; inappropriate responses to behavior.

C. Values and Supports Student Diversity

- ☐ Recognized and consistently responded to the diversity in the class (gender, ethnicity, academic and physical abilities); Consistently used or attempted to use strategies to address the needs of all students;
- ☐ Recognized but inconsistently responded to the student diversity; used or attempted to use some different strategies to address the needs of different students
- ☐ Little or no recognition or response to student diversity and individual needs; used the same approach for all students.

D. Fosters Mutual Respect Between Teacher and Students and Among Students

- ☐ Always treated all students with respect; encouraged and clearly expected students to treat each other with respect.
- ☐ Generally treated students with respect; some encouragement of students to treat each other with respect.
- ☐ Did not show respect or concern for students; little or no encouragement of students to treat each other with respect.

E. Provides a Safe Environment for Learning

- ☐ Classroom environment was emotionally and physically safe for students at all times.
- ☐ Classroom environment was emotionally and physically safe for students most of the time.
- ☐ Classroom environment was not emotionally and/or physically safe for students.

VIII. TEACHER IMPLEMENTS AND MANAGES INSTRUCTION (Mark one response for each section)

A. Implements Instruction Based on Student Needs and Assessment Data

- ☐ Instruction addressed individual student needs; always used or attempted to use an appropriate instructional strategy to meet individual student needs; adapted instruction to changing or unanticipated circumstances.
- ☐ Instruction addressed most individual student needs; used more than one strategy as needed; sometimes adapted instruction to meet changing or unanticipated circumstances.
- ☐ Instruction did not address individual student needs; one strategy was used for all students; no attempt to adapt lesson to meet changing or unanticipated circumstances.

B. Uses Time Effectively

- ☐ Always used efficient procedures for non-instructional tasks (handling materials/supplies, managing transitions, organizing work, etc.) so there is minimal loss of learning time.
- ☐ Inconsistently used efficient procedures for non-instructional tasks causing some loss of learning time.
- ☐ Used inefficient procedures for non-instructional tasks resulting in significant loss of learning time.

C. Uses Space and Materials Effectively

- ☐ Consistently used classroom space and materials effectively to facilitate student learning.
- ☐ Classroom space and/or materials were not always used effectively to facilitate student learning.
- ☐ Ineffective use of classroom space and materials to facilitate student learning.

D. Implements and Manages Instruction to Facilitate Higher Order Thinking

- ☐ Most instruction encouraged higher order thinking of all students.
- ☐ Some instruction encouraged some higher order thinking by most students.
- ☐ Little or no instruction encouraged higher order thinking by any students.

IX. TEACHER ASSESSES AND COMMUNICATES LEARNING RESULTS (Mark one Response For Each Section)

A. Uses Formative Assessments Aligned with Learning Objectives

- ☐ Formative assessment strategies fully aligned with learning objectives; obviously used to adjust instruction.
- ☐ Formative assessment strategies aligned with learning objectives; appeared to be used to adjust instruction.
- ☐ Formative assessment strategies were generally aligned with learning objectives; not clear if or how used to adjust instruction.
- ☐ Formative assessment to support student learning not clearly aligned with objectives; appeared to be done without intention or done for compliance.
- ☐ No assessment strategies used even though formative assessment was needed to determine level of student learning.

B. Uses a Variety of Formative and/or Summative Assessments to Measure Student Learning

- ☐ Used assessment strategies which provided all students several opportunities to demonstrate learning.
- ☐ Used assessment strategies which provided most students opportunities to demonstrate learning.
- ☐ Used some assessment stategies which provided some students opportunities to demonstrate learning.
- ☐ Limited use of assessment strategies which provided minimal opportunities for students to demonstrate learning.
- ☐ No assessment strategies used even though formative assessment was needed to determine level of student learning.

C. Adapts Formative and/or Summative Assessments to Accommodate Diverse Learning Needs and Situations.

- ☐ Assessment strategies were obviously adapted to accommodate student diversity and diverse learning needs.
- ☐ Assessment strategies appeared to be adapted to accommodate student diversity and diverse learning needs.
- ☐ Some attempts to adapt assessment strategies to meet diverse needs however not successful for all students.
- ☐ Limited attempt to adapt assessment strategies to accommodate student diversity or diverse student needs.
- ☐ No assessment strategies used even though formative assessment was needed to determine level of student learning.

X. OVERALL CLASSROOM RATING PROFILE (Mark only one)

- ☐ Instruction was effective for all students; evidence that instruction based on clearly defined objectives fully aligned with standards; all students engaged in activities requiring higher level thinking skills.
- ☐ Instruction was effective for most students; evidence that instruction based on clearly defined objectives aligned with standards; most students engaged in activities which required higher level thinking skills.
- ☐ Instruction was somewhat effective for most students; evidence that instruction was based on student objectives somewhat aligned with standards; some opportunity for higher level thinking skills development.
- ☐ Instruction observed was of poor-mediocre quality and effective for only a portion of the students; little evidence that instruction was based on student objectives; instruction had a minimal impact on learning.
- ☐ Instruction was of poor quality and was not effective for any students; no evidence that instruction was based on student objectives; learning was not based on instruction provided.

TO IDENTIFY INSTRUCTIONAL ENVIRONMENT CONTEXT ONLY

PHYSICAL SETTING/CLASSROOM ENVIRONMENT (Mark all that Apply in sections A, B, C)

A. Classroom Facilitates Student Learning	C. Classroom Environment
☐ Student seating is flexible to allow for differing needs (projects, experimentation, cooperative groups, etc.)	☐ Science Materials/Equipment evident ☐ Science displays promote learning
☐ Needed utilities are available (water, electricity, etc.)	☐ Science reference books available
☐ Flat top surfaces are sufficient for experimentation, projects, displays, etc.	☐ Student textbooks evident ☐ Computers available for student use # ___
B. Classroom Facility	☐ Ongoing science projects in evidence
☐ Classroom adequate size for student number	☐ Student work displayed
☐ Adequate storage for resources/materials/equipment ☐ Furnishings allow for activity-based instruction	☐ Living materials present (according to school policy and when appropriate)

Mathematics classroom observation instrument

MATHEMATICS CLASSROOM OBSERVATION INSTRUMENT Code:
Leadership by Design (NBC Version)
Copyright: Briarwood Enterprises LLC.

Level/Class ___ **Lesson Title** _____ **Total # Students** ____ **Gender: M**____**F** ____
 # Minority ____ **# Inclusion** ____ **Length of Observation** _____

Learning Objective of the Lesson_____

I. LESSON OVERVIEW

A. Learning Objective of the Lesson ***(Mark all that apply)***

☐ Clearly communicated by the teacher using multiple means ☐ Communicated orally only ☐ Communicated in writing only ☐ Student activities consistent with the lesson objective(s) ☐ Student activities not consistent with the lesson objective(s) ☐ Lesson objective communicated but not clear ☐ Lesson objective not communicated

B. Major Instructional Resource used in the lesson observed ***(Mark 1, 2, 3…with 1 meaning"primary/predominant resource influencing instruction"))***

☐ Textbook ☐ Other Print Materials (worksheet, manual, etc.) ☐ Technology based presentation media (SMART Board, Power Point, etc.) ☐ Document Camera ☐ Manipulatives/Hands-on materials _____
☐ Calculators (Other than Graphing) ☐ Computer ☐ Graphing Calculator ☐ Mathematics Centers
☐ Technology Probes ☐ None of the above Other_____

C. 1. Content Delivery *(Mark all that apply)*	**D. Place in Instructional Sequence** *(Mark 1, 2, 3…)*	**E. Seating Arrangement for Lesson** *(Mark 1, 2, 3…)*
☐ Age/grade level appropriate	☐ Introduction of new concept	☐ Whole group
☐ Content presented is accurate	☐ Develop conceptual understanding	☐ Small groups working on same task
☐ One or more content errors	☐ Apply concept to new situation	☐ Small groups working on different task
☐ Student misconception not Corrected	☐ Review concept or procedure	☐ Individuals working on same task
	☐ Assess student understanding	☐ Individuals working on different tasks

C. 2. Content Focus ***(Mark 1, 2, 3…)*** ☐ Number/Computation ☐ Geometry ☐ Measurement ☐ Probability ☐ Statistics ☐ Algebra ☐ Pre-calculus/Calculus

II. INSTRUCTIONAL OVERVIEW ***(Mark 1, 2, 3… in each section with 1 meaning "primary/predominant resource influencing instruction")***

A. Instructional Strategy

☐ Teacher lecture ☐ Teacher demonstration ☐ Teacher-led discussion ☐ Individual assistance
☐ Student presentation ☐ Small group discussion ☐ Students Solving Problems ☐ Other _____

B. Student Activity

☐ Listening to/observing teacher presentation ☐ Participating in discussion (teacher led or small group)
☐ Conducting mathematics investigation ☐ Completing a skills/practice worksheet (recall or comprehension)
☐ Higher-level problem-solving assignment ☐ Using hands-on materials to solve problems/ verify solutions
☐ Applying math to realistic problems ☐ Assignment/answering questions from text/other resources
☐ Taking test ☐ Sharing solutions or strategies ☐ Using computer software program ☐ Using the Internet for research
☐ Using computer for inputting/analyzing data

Comments:

III. QUESTIONING

A. Quality of Questions ***(Mark ONLY ONE box, record examples of each)***

☐ Questions were mostly narrow or convergent focusing on factual recall or one word responses
☐ Questions were mostly broad or divergent and stimulated higher cognitive student responses
☐ Appropriate balance of factual recall and higher cognitive questions
☐ No questions were asked by teacher or posed through the activity being conducted

B. Questioning Techniques ***(Mark all that Apply)***

☐ Students are encouraged to ask questions of each other and/or the teacher ☐ Questions stimulated higher level and divergent thinking ☐ Appropriate wait time ☐ **All** students have an opportunity to respond* ☐ **Most** students have an opportunity to respond* ☐ **Only a few** students have an opportunity to respond* ☐ Teacher provides focused, descriptive, and qualitative feedback to student responses*

IV. CLASSROOM ATMOSPHERE

A. Student Involvement *(Mark only one)*

- ☐ All or nearly all students demonstrate interest and were engaged
- ☐ Majority of students demonstrate interest, were engaged
- ☐ Approximately equal numbers of students interested/engaged and not interested/not engaged
- ☐ Majority of students uninterested or apathetic; generally not engaged
- ☐ Nearly all of the students were uninterested and not engaged

B. Classroom Management *(Mark only one)*

- ☐ Classroom orderly, no student disruptions (or minor) that impaired learning environment
- ☐ Classroom generally orderly but some student disruptions that required disciplinary action
- ☐ Classroom disorderly, frequent student disruptions that seriously impaired the learning environment

C. Classroom Culture/Learner Attitudes Demonstrated *(Mark all that apply)*

- ☐ Curiosity ☐ Cooperation with teacher and/or other students ☐ Persistence ☐ Responsibility
- ☐ Confidence in ability to "do" math ☐ Enthusiasm for learning math ☐ Accuracy ☐ Use of critical thinking skills

V. ANALYSIS OF INSTRUCTION LEADING TO THE DEVELOPMENT OF HIGHER ORDER SKILLS

A. Amount of Problem Solving/Student Investigation/Research ***(Mark ONLY ONE)***

- ☐ Students are engaged in a mathematics problem solving/inquiry experience which may include skills 1-7, however the emphasis is on higher level skills 8-17.
- ☐ Students are engaged in problem/inquiry based activity in which the focus of lesson is on the lower level skills 1-7.
- ☐ Students are not involved in any type of problem-solving/inquiry/investigative activity. *(If marked, also mark B, 4th box)*

B. Level of Student Engagement in Problem Solving/Investigation/Research ***(Refers back to Part A)(Mark ONLY ONE)***

- ☐ Students solve meaningful mathematical or realistic problems through explorations or investigations that can be generalized to allow them to make valid conjectures (#14), determine strategies to solve problems (#13), evaluate logical consistency (#15) and/or justify/verify solutions (#16).
- ☐ Students discover a mathematics phenomenon using a planned activity that requires using a problem-solving strategy, collecting and analyzing data, and/or making connections between mathematics ideas or strands.
- ☐ Students learn a mathematics concept using a preplanned activity that provides a definitive procedure and requires a specific response to be correct.
- ☐ Students are not involved in any type of problem solving/inquiry/investigative activity.

C. Mathematical Skills Being Developed ***(Mark all skills which are introduced/developed in the observed lesson)***

Basic Skills: ***(Mark all that are observed)***

- ☐ 1.Recognizing/observing ☐ 2. Reciting/recalling facts ☐ 3. Classifying ☐ 4. Measuring/estimating
- ☐ 5. Coordinate Graphing ☐ 6. Constructing charts/graphs ☐ 7. Computing/calculating

Higher Level Mathematical Skills: ***(Mark all that are observed)***

- ☐ 8. Collecting/recording data ☐ 9. Interpreting/analyzing data/statistics ☐ 10. Investigating (Hands-on, Tech)
- ☐ 11.Applying Theorems/principles ☐ 12. Evaluating the Relevancy of data
- ☐ 13. Determining problem solving strategy ☐ 14. Creating/formulating pattern or equation
- ☐ 15. Evaluating logical consistency ☐ 16. Justifying/verifying solutions ☐ 17. Interpretive Discussion

VI. TEACHER DEMONSTRATES APPLIED CONTENT KNOWLEDGE *(Mark one response for each section)*

A. Communication

- ☐ Consistently used accurate and effective communication; vocabulary is clear, correct and appropriate.
- ☐ Generally used accurate and effective communication; occasional use of inappropriate vocabulary.
- ☐ Consistently used inaccurate and ineffective communication and/or inappropriate vocabulary.

B. Connects Content to Life Experiences

- ☐ Consistently connected most content/procedures/activities with relevant life experiences.
- ☐ Connected some content/procedures/activities with relevant life experiences.
- ☐ Rarely or never connected content/procedures/activities with relevant life experiences.

C. Instructional Strategies Appropriate for Content and Contribute to Student Learning

- ☐ Used instructional strategies that were clearly appropriate for the content/processes of the lesson.
- ☐ Used instructional strategies that were generally appropriate for the content/processes of the lesson.
- ☐ Used instructional strategies that were questionable or inappropriate for the content/processes of the lesson.

D. Guides Students to Understand Lesson Content from Various Perspectives to Extend Understanding

- ☐ Provided multiple opportunities for students to consider content from a different context or perspective.
- ☐ Provided a single opportunity for students to consider content from a different context or perspective.
- ☐ Never provided an opportunity for students to consider content from different context or perspective.

VII. TEACHER CREATES AND MAINTAINS LEARNING CLIMATE *(Mark one response for each section)*

A. Communicates High Expectations

- ☐ Significant/challenging lesson objectives; teacher consistently communicates confidence in students' ability to achieve.
- ☐ Challenging objectives; some communication of confidence in students' ability to achieve.
- ☐ Minimal objectives for students; rarely or never communicates confidence in students' ability to achieve.

B. Establishes a Positive Learning Environment

- ☐ Clear conduct standards; awareness of student behavior; responded appropriately/respectfully.
- ☐ Conduct standards but some inconsistency in monitoring and response to student behavior.
- ☐ No established conduct expectations; minimal or no monitoring; inappropriate responses to behavior.

C. Values and Supports Student Diversity

- ☐ Recognized and consistently responded to the diversity in the class (gender, ethnicity, academic and physical abilities); Consistently used or attempted to use strategies to address the needs of all students;
- ☐ Recognized but inconsistently responded to the student diversity; used or attempted to use some different strategies to address the needs of different students
- ☐ Little or no recognition or response to student diversity and individual needs; used the same approach for all students.

D. Fosters Mutual Respect Between Teacher and Students and Among Students

- ☐ Always treated all students with respect; encouraged and clearly expected students to treat each other with respect.
- ☐ Generally treated students with respect; some encouragement of students to treat each other with respect.
- ☐ Did not show respect or concern for students; little or no encouragement of students to treat each other with respect.

E. Provides a Safe Environment for Learning

- ☐ Classroom environment was emotionally and physically safe for students at all times.
- ☐ Classroom environment was emotionally and physically safe for students most of the time.
- ☐ Classroom environment was not emotionally and/or physically safe for students.

VIII. TEACHER IMPLEMENTS AND MANAGES INSTRUCTION *(Mark one response for each section)*

A. Implements Instruction Based on Student Needs and Assessment Data

- ☐ Instruction addressed individual student needs; always used or attempted to use an appropriate instructional strategy to meet individual student needs; adapted instruction to changing or unanticipated circumstances.
- ☐ Instruction addressed most individual student needs; used more than one strategy as needed; sometimes adapted instruction to meet changing or unanticipated circumstances.
- ☐ Instruction did not address individual student needs; one strategy was used for all students; no attempt to adapt lesson to meet changing or unanticipated circumstances.

B. Uses Time Effectively

- ☐ Always used efficient procedures for non-instructional tasks (handling materials/supplies, managing transitions, organizing work, etc.) so there is minimal loss of learning time.
- ☐ Inconsistently used efficient procedures for non-instructional tasks causing some loss of learning time.
- ☐ Used inefficient procedures for non-instructional tasks resulting in significant loss of learning time.

C. Uses Space and Materials Effectively

- ☐ Consistently used classroom space and materials effectively to facilitate student learning.
- ☐ Classroom space and/or materials were not always used effectively to facilitate student learning.
- ☐ Ineffective use of classroom space and materials to facilitate student learning.

D. Implements and Manages Instruction to Facilitate Higher Order Thinking

- ☐ Most instruction encouraged higher order thinking of all students.
- ☐ Some instruction encouraged some higher order thinking by most students.
- ☐ Little or no instruction encouraged higher order thinking by any students.

IX. TEACHER ASSESSES AND COMMUNICATES LEARNING RESULTS *(Mark one response each section)*

A. Uses Formative Assessments Aligned with Learning Objectives

- ☐ Formative assessment strategies fully aligned with learning objectives; obviously used to adjust instruction.
- ☐ Formative assessment strategies aligned with learning objectives; appeared to be used to adjust iinstruction.
- ☐ Formative assessment strategies were generally aligned with learning objectives; not clear if or how used to adjust instruction.
- ☐ Formative assessment to support student learning not clearly aligned with objectives; appeared to be done without intention or done for compliance.
- ☐ No assessment strategies used even though formative assessment was needed to determine level of student learning.

B. Uses a Variety of Formative and/or Summative Assessments to Measure Student Learning

- ☐ Used assessment strategies which provided all students several opportunities to demonstrate learning.
- ☐ Used assessment strategies which provided most students opportunities to demonstrate learning.
- ☐ Used some assessment strategies which provided some students opportunities to demonstrate learning.
- ☐ Limited use of assessment strategies which provided minimal opportunities for students to demonstrate learning.
- ☐ No assessment strategies used even though formative assessment was needed to determine level of student learning.

C. Adapts Formative and/or Summative Assessments to Accommodate Diverse Learning Needs and Situations.

- ☐ Assessment strategies were obviously adapted to accommodate student diversity and diverse learning needs.
- ☐ Assessment strategies appeared to be adapted to accommodate student diversity and diverse learning needs.
- ☐ Some attempts to adapt assessment strategies to meet diverse needs however not successful for all students.
- ☐ Limited attempt to adapt assessment strategies to accommodate student diversity or diverse student needs.
- ☐ No assessment strategies used even though formative assessment was needed to determine level of student learning.

X. OVERALL CLASS INSTRUCTION RATING PROFILE *(Mark only one)*

- ☐ Instruction was effective for all students; evidence that instruction based on clearly defined objectives fully aligned with standards; all students engaged in activities requiring higher level thinking skills.
- ☐ Instruction was effective for most students; evidence that instruction based on clearly defined objectives aligned with standards; most students engaged in activity that required or offered opportunity to develop higher level thinking skills.
- ☐ Instruction was somewhat effective for most students; evidence that instruction was based on student objectives somewhat aligned with standards; little opportunity for higher level thinking skills development.
- ☐ Instruction observed was of mediocre quality and effective for only a portion of the students; little evidence that instruction was based on student objectives; instruction had a minimal impact on learning.
- ☐ Instruction was of poor quality and was not effective for any students; no evidence that instruction was based on student objectives; learning was not based on instruction provided.

PHYSICAL SETTING/CLASSROOM ENVIRONMENT *(Mark all that apply in sections A, B, C)*

A. Classroom Facilitates Student Learning	C. Classroom Environment
☐ Student seating is flexible to allow for differing needs (individual work, cooperative groups, etc.)	☐ Mathematics manipulatives/tools evident*
	☐ Mathematics displays promote learning*
☐ Adequate electrical and any other needed utilities available	☐ Class sets of calculators available*
☐ Flat top surfaces are sufficient for working with hands-on materials or problems, projects, models, displays, etc.	☐ Mathematics textbooks evident*
	☐ Computers (student) available, # _____
B. Classroom Facility	☐ Ongoing mathematics projects in evidence
☐ Classroom adequate size for student number	☐ Mathematics student work displayed
☐ Adequate storage for resources/materials/equipment	☐ Adequate resources for lesson are present
☐ Furnishings allow for activity-based instruction (CBI., etc)	☐ Outside interruptions (#_____)

Appendix B: Rubric for scoring classroom observations

Science rubric

Scoring Rubric – Science Classroom Observations

Scores of 5-1 reflect the perceived status of instruction; NO = Not observed but could have contributed to the lesson; NA = Not observed and not applicable for meeting lesson objective

LBD Observation Section	Score of 5	Score of 4	Score of 3	Score of 2	Score of 1
I. Lesson Overview					
A. Lesson Objectives **Score**	Objectives for the lesson were clear, appropriate and communicated in multiple ways; student activities were totally consistent with the communicated lesson objectives; lesson targets were appropriate and clearly defined so all students understood them.	Objectives for the lesson were clear, appropriate and communicated in at least one way; student activities were consistent with the communicated lesson objectives; lesson targets were appropriate and defined so that most students understood them.	Objectives for the lesson were appropriate but not fully communicated and not readily apparent to the students; student activities were generally consistent with the perceived lesson objectives; lesson targets were appropriate although not fully defined so that all students understood them.	Objectives for the lesson may be appropriate but were not communicated in any way to the students; student activities were only partially consistent with the perceived lesson objectives; lesson targets were not fully defined and only a few of the students seemed to understand them.	No particular objective was evident for the lesson or the objective has no connection to the activity; lesson targets were not defined so that any of the students understood them.
B. Use of Instructional Resources **Score**	Instructional resources were appropriate for the activity, well designed, and fully consistent with lesson objectives; were suitable for and of interest to all students.	Instructional resources were appropriate for the activity, well designed, and consistent with lesson objectives; resources were suitable for and of interest to nearly all of the students.	Instructional resources were appropriate for the activity but not totally consistent with the lesson objectives; resources were suitable for and of interest to half or more of the students.	Instructional resources were appropriate for the activity but other, more effective resources are available and more consistent with the lesson objectives; resources were suitable for and of interest to only a few students.	Instructional resources were not appropriate for the activity and did not assist student learning.
C. Content Delivery **Score**	The lesson was well designed to achieve the lesson objectives; appropriate highly effective instructional practices were used at all times to achieve the lesson objectives.	The lesson was well designed to achieve the lesson objectives; appropriate highly effective practices were used to achieve the lesson objectives most of the time.	The lesson was adequately designed to achieve the lesson objectives; the practices used were appropriate, however other practices might have been more effective.	The lesson design was not clear or not fully consistent with the lesson objectives; The practices used were appropriate; however other practices would have been more effective.	The lesson was not designed to achieve any lesson objective; the instruction did not use appropriate practices.

D. Place in Instructional Sequence **Score**	The lesson was well designed and fully fits with its place in the instructional sequence; the lesson contributed to the achievement of the stated overall learning objectives.	The lesson was adequately designed for its place in the instructional sequence; the lesson generally contributed to the achievement of the stated overall learning objectives.	The lesson was adequately designed and generally consistent with its phase in the instructional sequence; the lesson only partially contributes to the achievement of the overall learning objectives.	The lesson was adequately designed but not fully consistent with its phase in the instructional sequence; lesson contributes minimally to the achievement of the overall learning objectives.	The lesson was of poor design and inconsistent with its phase in the instructional sequence; the lesson did not contribute to the achievement of overall learning objectives.
E. Seating Arrangement for Lesson **Score**	Students were seated in an appropriate configuration that optimized effective student learning and was fully consistent with and contributed to the lesson's instructional objectives.	Students were seated in an appropriate configuration that enabled effective student learning; arrangement was consistent with the lesson's instructional objectives.	Students were seated in a configuration that was appropriate but may not have been the best for achieving the lesson's instructional objectives; generally consistent with the lesson's objectives.	Students were seated in a configuration that was appropriate but was not suited for achieving the lesson's instructional objectives; seating was inconsistent with the lesson's objectives.	The student seating configuration was not appropriate and not conducive to accomplishing the lesson's instructional objectives.
Section I Comments					
II. Instructional Overview					
A. Student Focus **Score**	Instruction was student-centered and all, or nearly all, students took responsibility to fully participate in the work through discussion and creation of appropriate products/utilization of science skills; the instructional activity was engaging for nearly all or all of the students.	Instruction was student-centered and most of the students took responsibility to fully participate in the work through discussion and creation of appropriate products/utilization of science skills; the instructional activity was engaging for most of the students.	Instruction was student-centered and approximately equal numbers of students fully participate in the work through discussion and creation of appropriate products/utilization of science skills; the instructional activity was engaging for some, but not all, of the students.	Instruction was mostly teacher directed; fewer than half of the students fully participated in the work through discussion and creation of appropriate products/ utilization of science skills; the instructional activity was engaging for only a few of the students.	Instruction was totally teacher directed, not at all student-centered and was not engaging for the students; students may have been compliant, but few if any were clearly engaged in the instructional activity.
B. Instructional strategies	Instruction was varied, included students in presenting or discussion, and incorporated activity-	Instruction was varied and incorporated activity-based and/or technology resources as appropriate and needed;	Instruction included only one or two strategies but incorporated appropriate activity-based and/or	Instruction incorporated some activity-based or technology resources; the strategy used did not result	Instruction incorporated few if any activity-based or technology resources; the strategy used did not

Score	based and/or technology resources as appropriate and needed; the resources used were fully effective in reaching the lesson's objectives for all students.	the resources used were generally effective in reaching the lesson's objectives for most students.	technology resources; however the resources used were not fully effective in reaching the lesson's objectives for some students.	in student learning for many students; a different resource would have been more appropriate.	seem to result in student learning for any students; a different resource was needed.
C. Awareness of student needs **Score**	Instructional strategies reflected current understanding about the way children learn; teacher always utilized appropriate interventions; the teacher differentiated instruction to meet the needs of individual students.	Instructional strategies reflected a general understanding about the way children learn; teacher utilized appropriate interventions; teacher sometimes differentiated instruction to meet the needs of individual students.	Instructional strategies reflected a general understanding of the way children learn; teacher utilized some appropriate interventions; some effort to differentiate instruction to meet the needs of individual students.	Instructional strategies reflected a minimal understanding of the way children learn; teacher occasionally utilized appropriate interventions; minimal effort to differentiate instruction to meet the needs of individual students	Instructional strategies did not reflect an understanding of the way children learn; teacher did not utilize interventions or differentiate instruction to meet the needs of individual students
Section II Comments					
III. Questioning					
A. Quality of the Questions **Score**	Many significant questions were posed which stimulated broad student responses; most questions were divergent and required higher level thinking of all students.	Several significant questions were posed which stimulated broad student responses; an appropriate balance of divergent and convergent questions many of which required higher level thinking.	A few significant questions were posed and, although some questions were divergent and stimulated broad student responses, the majority of the questions were convergent and focused on factual recall.	Few, if any, significant questions were asked which stimulated broad student responses; questions were all or nearly all convergent, focusing on factual recall.	No questions were asked or if asked, were all convergent and did not elicit many student responses; all questions were focused on factual recall.
B. Participation in Questioning and Discussion **Score**	All students had an opportunity to respond and recognized that they may be expected to share at any time. Both the teacher and students initiated significant questions and all students were consistently encouraged to ask questions of each other.	Most students had an opportunity to respond and recognized that they may be expected to share again. Both the teacher and students initiated questions and many students had opportunities for students to ask questions of each other.	Many students had an opportunity to respond, however, once they responded, there appeared to be little chance they would be called on again. Several students generated questions, but most were initiated by the teacher. Students had limited opportunities to ask questions of each other.	A few students, generally raising their hands, had an opportunity to respond to the questions initiated by the teacher. There was little or no encouragement for students to ask questions of each other and there were few, if any, student questions.	There were no or very limited opportunities for students to respond to questions. All questions were initiated by the teacher. Students did not have an opportunity or any encouragement to ask questions of each other.

C. Target-centered Questions **Score**	The use of strategic or target-centered questions for formative assessment was intentional and clearly planned; all student responses were considered and used to adjust the pace and focus of the lesson.	The use of some target-centered questions for formative assessment was intentional, with some evidence of planning for them; many, but not all, student responses were considered and/or utilized to alter the pace or focus of the lesson.	The use of some target-centered questions for formative assessment was intentional but with little evidence of planning for them; many, but not all, student responses were considered and/or utilized to alter the pace or focus of the lesson.	The use of target-centered questions for formative assessment was limited, with no evidence of planning for them; some student responses were considered and/or utilized to alter the pace or focus of the lesson.	The intentional use of target-centered questions for formative assessment was not evident; responses by students did not alter the pace or focus of the lesson.
D. Feedback to Responses **Score**	Students always had ample time to consider the question before responding; appropriate feedback (e.g., clear, specific, and descriptive) was consistently given to all students; feedback always encouraged student involvement in the discussion/task.	Students usually had ample time to consider the question before responding; appropriate feedback (e.g., clear, specific, and descriptive) was given to nearly all the students; feedback usually encouraged student involvement in the discussion/task.	Students generally had time to consider the question before responding however, response time varied and/or was not consistent; appropriate feedback (e.g., clear, specific, and descriptive) was given to many of the students; feedback did not always encourage student involvement.	Students had minimal time to consider the question before responding; appropriate feedback (e.g., clear, specific, and descriptive) was given occasionally or seemed to be given only to a few students; limited encouragement for students to become involved.	Students had no time to consider the question before responding; appropriate feedback (e.g., clear, specific, and descriptive) was not given or given only to a few students; no encouragement for student involvement.
Section III Comments					
IV. Classroom Atmosphere					
A. Student Involvement **Score**	All of the students demonstrated interest and were engaged in the instructional activity.	Most of the students demonstrated interest and were engaged in the instructional activity.	Approximately equal numbers of students were interested/engaged and not interested/engaged in the instructional activity.	Only a few of the students were interested/engaged in the instructional activity.	None of the students were interested/engaged in the instructional activity.
B. Classroom Management **Score**	The classroom was well managed and totally orderly, no student disruptions that caused a loss of instructional time or impaired the learning environment.	The classroom was well managed and orderly; some minor student incidents that did not require any corrective action and did not cause any loss of instructional time.	The classroom was generally well managed and orderly; there were one or a few minor student disruptions that required corrective or disciplinary action and caused a minimal loss of instructional time.	The classroom was poorly managed and/or disorderly with frequent student disruptions that required corrective or disciplinary action and caused a significant loss of instructional time.	The classroom was disorderly; constant student disruptions that caused a major loss of instructional time and seriously impaired the learning environment.

C. Classroom culture **Score**	The teacher has established a classroom culture in which all, or nearly all, of the students took initiative in discussions and activities; all students demonstrated respect for other students; all, or nearly all demonstrated a willingness to express alternative views.	The teacher has established a classroom culture in which most of the students took initiative in discussions and activities; most students demonstrated a respect for other students; most students demonstrated a willingness to express alternative views.	The teacher has established a classroom culture in which many (majority) of the students took initiative in discussions and activities; most students demonstrated respect for other students; many students demonstrated a willingness to express alternative views.	The teacher has established a classroom culture in which only a few students took initiative in discussions and activities; the majority demonstrated a respect for other students; only a few students demonstrated a willingness to express alternative views.	The teacher has established a classroom culture in which none of the students took initiative in discussions and activities; some students demonstrated a respect for other students; no or only a few, students demonstrated a willingness to express alternative views.
Section IV Comments					
V. Analysis of Instruction Leading to the development of Higher Order Science Skills					
A. Amount/Level of Student Investigation **Score**	Students were engaged in an investigation/ experimentation that utilized higher level thinking skills. Students designed and carried out an experiment to solve a problem initiated by a teacher or student question.	Students were engaged in an investigation/experimentation that utilized higher level thinking skills. Students carried out an investigation/ experiment designed/ selected by the teacher to teach a particular concept.	Students were engaged in investigation/ experimentation that used some higher-level skills, however the focus of lesson was on the basic skills. Students used a teacher-developed activity that required the collection and analysis of data (beyond merely reporting results) to solve a problem or create a product.	Students were not involved in any type of investigation or were involved in an investigation that focused on lower level science skills; in some cases, the lesson objective did not require an investigative activity would have been enhanced by it.	Students were not involved in any type of investigation/or experimentation but the lesson clearly needed it; there was no evidence the content/concept was learned using the strategies employed in this lesson.
B. Scientific skills being developed **Score**	Students were developing/utilizing higher-level scientific skills during the entire class period; interpretative discussions involved all students at a high-level.	Students were developing/ utilizing higher-level scientific skills during most of the class period; interpretative discussions involved most of the students at a high-level.	Students were observed utilizing higher-level skills; however, the focus of the investigation or experiment was on lower level skills during most of the period; minimal interpretative discussion or interpretive discussion that engaged only part of the students.	Students were not observed utilizing higher level skills and, if using lower level skills, the investigation or experiment did not require data collection or analysis; little or no interpretive discussion.	Students were not involved in any type of investigation/or experimentation; no development of basic or higher level scientific skills; no interpretive discussion.
Section V Comments					

VI. The Teacher Demonstrates Applied Content Knowledge					
A. Communication **Score**	Consistently used accurate and effective communication; vocabulary was clear, correct and appropriate.		Generally used accurate and effective communication; occasional use of inappropriate vocabulary; exhibited some minor errors that do not interfere with conceptual development		Consistently used inaccurate, misleading and/or ineffective communication and/or inappropriate vocabulary.
B. Connects Content to Life Experiences **Score**	Consistently connected content, procedures, and activities with relevant life experiences, current events and/or significant historical events.		Connected some content, procedures, and activities with relevant life experiences, current events and/or significant historical events.		Rarely/never connected content, procedures or Activities with relevant life experiences, current events, and/or significant historical events.
C. Instructional Strategies appropriate for content and contribute to student Learning **Score**	Used instructional strategies that were clearly appropriate for the content/processes of the lesson; evident that student learning occurred as a result of the strategies employed.		Used instructional strategies that were generally appropriate for the content/processes of the lesson; however, not clear if the student learning was a result of the strategies employed.		Used instructional strategies that were questionable or inappropriate for the content/processes of the lesson; no indication that student learning occurred.
D. Guides Students to Understand Lesson Content from Various Perspectives **Score**	Provided multiple opportunities for students to consider content from different perspectives or contexts.		Provided one or few opportunities for students to consider content from different perspectives or contexts.		Rarely/never provided opportunities for students to consider content from different perspectives or contexts.
Section VI Comments:					
VII. The Teacher Creates and Maintains a Positive Learning Climate					
A. Communicates High Expectations **Score**	Presented significant and challenging objectives; consistently communicated confidence in students' ability to achieve.		Presented challenging objectives; and at times, communicated confidence in students' ability to achieve.		Presented minimal or no objectives for students; rarely or never communicated confidence in students' ability to achieve.

B. Establishes a Positive Learning Environment **Score**	Clear conduct/safety standards have been established and are being met; awareness of student behavior; responded appropriately/ respectfully to student misbehavior.		Conduct/safety standards have been established; there was some inconsistency in monitoring and response to student misbehavior.		No established conduct/safety standards or expectations; minimal or no monitoring; inappropriate responses to student misbehavior.
C. Values and Supports Student Diversity (including ethnicity, gender, socioeconomic status) **Score**	Recognized and consistently responded to the diversity in the class; consistently used or attempted to use strategies to address the needs of all students.		Recognized but inconsistently responded to the diversity in the class; used or attempted to use some different strategies to address the needs of particular students.		Provided little or no recognition or response to student diversity and individual needs; used the same approach for all students.
D. Fosters Mutual Respect Between Teacher and Students and Among Students **Score**	Always treated all students with respect; encouraged and clearly expected students to treat each other with respect.		Generally treated students with respect; provided some encouragement of students to treat each other with respect.		No evidence of the teacher's respect or concern for students was observed; provided little or no encouragement of students to treat each other with respect.
E. Provides a Safe Environment for Learning **Score**	Classroom environment was emotionally and physically safe for students at all times.		Classroom environment was emotionally and physically safe for students most of the time.		Classroom environment was not emotionally and/or physically safe for students.
Section VII Comments:					

VIII. The Teacher Implements and Manages Instruction Leading to Positive Student Outcomes					
A. Implements Instruction Based on Student Needs and Assessment Data **Score**	Instruction addressed all individual student needs; always used or attempted to use a variety of appropriate instructional strategies to meet individual student needs; adapted instruction to changing or unanticipated circumstances	Instruction addressed most individual student needs; used different instructional strategies as needed to meet needs of most students; sometimes adapted instruction to meet changing or unanticipated circumstances	Instruction addressed many individual student needs; used more than one instructional strategy as needed; occasionally adapted instruction to meet changing or unanticipated circumstances	Instruction addressed some individual student needs; attempted to use more than one instructional strategy; seldom adapted instruction to meet changing or unanticipated circumstances	Instruction did not address individual student needs; one strategy was used for all students; no attempt to adapt lesson to meet changing or unanticipated circumstances.
B. Uses Time, Space, and Materials Effectively **Score**	Always used efficient procedures for non-instructional tasks; no loss of learning time was observed; classroom space and materials were always used effectively to facilitate student learning.	Used efficient procedures for non-instructional tasks most of the time; minimal loss of learning time was observed; classroom space and materials were used effectively to facilitate student learning.	Generally used efficient procedures for non-instructional tasks with some loss of learning time; classroom space and materials were used effectively most of the time.	Used both efficient and inefficient procedures for non-instructional tasks resulting in significant loss of learning time; classroom space and/or materials were used effectively to facilitate student learning some of the time.	Used inefficient procedures for non-instructional tasks resulting in major loss of learning time; classroom space and materials were not used effectively to facilitate student learning.
C. Implements and Manages Instruction to Facilitate Higher Order Thinking **Score**	Instruction encouraged higher order thinking by all students; included significant amount of independent and/or group processing and reflection time.	Instruction encouraged higher order thinking by most students; included some independent or group processing and reflection time.	Instruction encouraged higher order thinking by some students; included minimal independent or group processing and reflection time.	Instruction encouraged higher order thinking by only a few students; little, if any, independent or group processing or reflection time was provided.	Instruction was minimal and ineffective; did not encourage higher order thinking by any students; did not include any independent/group processing or reflection time
Section VIII Comments					
IX. The Teacher Assesses and Communicates Learning Results					
A. Uses Assessments Aligned with Learning Objectives	Formative assessment strategies were fully aligned with learning objectives; assessment results were obviously	Formative assessment strategies were aligned with learning objectives; appeared to be used to adjust instruction.	Formative assessment strategies were generally aligned with learning objectives; not clear if or how assessment results	Formative assessment strategies not clearly aligned with learning objectives; appeared to be done without intention or	No assessment strategies were used even though formative assessment was needed to determine the level of student
Score	used to adjust instructional practice in a timely manner.		were used to adjust instructional practice.	done for compliance only.	learning.
B. Uses a Variety of Formative and/ or Summative Assessments to **Score**	Used various formative and/or summative assessment strategies that provided all students several opportunities to demonstrate learning.	Used various formative and/or summative assessment strategies that provided most students some opportunities to demonstrate learning.	Used some formative and/or summative assessment strategies that provided many students (at least the majority) opportunities to demonstrate learning.	Limited use of formative and/or summative assessment strategies that provided opportunities for some students to demonstrate learning.	No assessment strategies were used even though formative assessment was needed to determine the level of student learning.
C. Adapts All Assessments to Accommodate Diverse Learning Needs and Situations **Score**	Formative and/or summative assessment strategies were obviously adapted to accommodate student diversity and diverse learning needs.	Formative and/or summative assessment strategies appeared to be adapted to accommodate student diversity and diverse learning needs.	Some attempts were made to adapt formative and/or summative assessment strategies to meet diverse needs however; these were not successful for all students.	A limited attempt was made to adapt assessment strategies to accommodate student diversity or to meet diverse student needs.	No assessment strategies were used even though formative assessment was needed to determine the level of student learning.
Section IX Comments					
X. Overall Classroom Observation Rating					
Overall rating of quality of instruction **Score**	Instruction was of high quality and effective for all students; evidence that instruction was based on clearly defined objectives that were fully aligned with standards; all students were engaged in activities requiring higher level thinking skills.	Instruction was of high quality and effective for most students; evidence that instruction was based on clearly defined objectives that were aligned with standards; most students were engaged in activities that required higher level thinking skills.	Instruction was of good quality and effective for many students; instruction appeared to be based on student objectives somewhat aligned with standards; some students had an opportunity for higher level thinking skills development.	Instruction was of mediocre quality and effective for only a small portion of the students; little evidence that instruction was based on student objectives; instruction had minimal impact on student learning.	Instruction was of poor quality and was not effective for any students; no evidence that instruction was based on student objectives; learning was not based on instruction provided.
Section X Comments:					

Physical Setting – This information is included for contextual purposes only. These data are not included in the rubric ratings.			
A. Classroom facilitates student learning **Score**	Flexible student furnishings can accommodate any type of activity to provide for maximum science activity interactions.	Student furnishings are somewhat flexible and can accommodate interaction for most types of science instructional activities.	Student furnishings are not flexible and in many cases, limit the interactions needed for quality science instruction.
B. Classroom facility **Score**	Classroom is large with sufficient storage for supplies; materials are well organized for ease of access; classroom furnishings are appropriate for hands-on activities and materials were available for all students.	Classroom is generally adequate in size with some storage; materials were generally well organized for ease of access; classroom furnishings are generally appropriate for hands-on activities but may not accommodate all students at the same time.	Classroom is inadequate in size with little or no storage; little, if any organization of materials for ease of access; classroom furnishings are not conducive to hands-on instruction or can only accommodate small numbers.
C. Classroom Environment **Score**	Science materials and equipment are abundant and easily obtained; ongoing student projects are evident and student work is prominently displayed.	Science materials and equipment are available but not in sufficient quantities; some student projects and limited amount of student work displayed.	Science materials are absent or extremely limited; no evidence of student projects and no student work displayed.
Physical Setting Comments			

November 15, 2013

Mathematics rubric

Scoring Rubric – Mathematics Classroom Observations

Scores of 5-1 reflect the perceived status of instruction; Score of NO = Not observed but could have contributed to the lesson; Score of NA = Not observed and not appropriate for meeting the lesson objective

LBD Observation Section	Score of 5	Score of 4	Score of 3	Score of 2	Score of 1
I. Lesson Overview					
A. Lesson Objectives **Score**	Objectives for the lesson are clear, appropriate and communicated in multiple ways; student activities are totally consistent with the communicated lesson objectives; lesson targets were appropriate, clearly defined and all students understood them.	Objectives for the lesson are clear, appropriate and communicated in at least one way; student activities are consistent with the communicated lesson objectives; lesson targets were appropriate and defined so that most students understood them.	Objectives for the lesson are appropriate but not fully communicated and not readily apparent to the students; student activities generally consistent with the perceived lesson objectives; lesson targets were appropriate although not fully defined so that all students understood them.	Objectives for the lesson may be appropriate but not communicated in any way to the students; student activities only partially consistent with the perceived lesson objectives; lesson targets were not fully defined and only a few of the students seemed to understand them.	No particular objective for the lesson or, the objective has no connection to the activity; lesson targets not defined so that any of the students understood them.
B. Use of Instructional Resources **Score**	Instructional resources were appropriate for the activity, well designed, and fully consistent with lesson objectives; were suitable for and of interest to all students.	Instructional resources were appropriate for the activity, well designed, and consistent with lesson objectives; were suitable for and of interest to nearly all of the students.	Instructional resources were appropriate for the activity but not totally consistent with the lesson objectives; were suitable for and of interest to half or more of the students.	Instructional resources were appropriate for the activity but other, more effective resources are available and more consistent with lesson objectives; resources suitable for and of interest to only a few students.	Instructional resources were not appropriate for the activity and did not assist student learning.
C. Content Delivery **Score**	The content presented is completely accurate and age/grade-level appropriate; it is delivered within a lesson designed to purposefully discover the common student misconceptions in order to	The content presented is completely accurate and age/grade-level appropriate; it is delivered within a well-designed lesson that may allow the teacher to discover some student misconceptions and	The content presented is accurate and age/grade-level appropriate; the lesson was not designed for the teacher to discover student misconceptions; if noticed were not clarified enough for the student(s)	The content presented is accurate but may or may not be age/grade-level appropriate; student misconceptions are either not noticed or not addressed.	The content presented is not accurate and/or not age/grade-level appropriate; student misconceptions were not even noticed.

B. Instructional strategies **Score**	Instruction was varied, included students in presenting or discussion, and incorporated activity-based and/or technology resources as appropriate and needed; the resources used were fully effective in reaching the lesson's objectives for all students.	Instruction was varied and incorporated activity-based and/or technology resources as appropriate and needed; the resources used were generally effective in reaching the lesson's objectives for most students.	Instruction included only one or two strategies but incorporated appropriate activity-based and/or technology resources; however, the resources used were not fully effective in reaching the lesson's objectives for some students.	Instruction incorporated some activity-based or technology resources; the strategy used did not result in student learning for many students; a different resource would have been more appropriate.	Instruction incorporated few if any activity-based or technology resources; the strategy used did not seem to result in student learning for any students; a different resource was needed.
C. Awareness of student needs **Score**	Instructional strategies reflected current understanding about the way children learn; teacher always utilized appropriate interventions; differentiated instruction to meet the needs of individual students.	Instructional strategies reflected a general understanding about the way children learn; the teacher utilized appropriate interventions; usually differentiated instruction to meet the needs of individual students.	Instructional strategies reflected a general understanding of the way children learn; teacher utilized some appropriate interventions; made some effort to differentiate instruction to meet the needs of individual students.	Instructional strategies reflected a minimal understanding of the way children learn; teacher occasionally utilized appropriate interventions; minimal effort to differentiate instruction to meet the needs of individual students	Instructional strategies did not reflect an understanding of the way children learn; teacher did not utilize interventions or differentiate instruction to meet the needs of individual students
Section II Comments					
III. Questioning					
A. Quality of the Questions **Score**	Many significant questions were posed which stimulated broad student responses; most questions were divergent and required higher level thinking skills.	Several significant questions were posed which stimulated broad student responses; an appropriate balance of divergent and convergent questions.	A few significant questions were posed and, although some questions were divergent and stimulated broad student responses, the majority of the questions were convergent and focused on factual recall.	Few, if any, questions were asked which stimulated broad student responses. Nearly all of the questions were convergent focusing on factual recall.	No questions were asked or if asked, were all convergent and did not elicit many, if any, student responses. All of the questions were focused on factual recall and initiated by the teacher.
B. Participation in questioning and discussion	All students had an opportunity to respond and recognized they may be expected to share at any time in pair/group	Most students had an opportunity to respond and recognized that they may be expected to share again in pair/group discussion,	Many students had an opportunity to respond, however it appeared that once they responded, there was little chance they	A few students, generally raising their hands, had an opportunity to respond; some consistently called out	There were no or very few opportunities for students to respond to questions; if students were in group settings,

Score	discussion, through sharing a problem-solving strategy, performing a task, or presenting a solution to a problem. Both the teacher and students initiated significant questions. All students were encouraged to ask questions of each other.	through sharing a problem-solving strategy, performing a task, or presenting a solution to a problem. Both the teacher and students initiated questions with a few opportunities for students to ask questions of each other.	would be called on again in pair/group settings. Some students participated in the discussion and a few shared a strategy, performed a task, or presented a solution to a problem. A few students generated questions but most were initiated by the teacher. Students had limited opportunities to ask questions of each other.	answers aloud. In group settings, it did not appear that students were expected to actively participate; a few shared out but many students worked individually. Questions were generally initiated by the teacher with few, if any, student questions.	they did not discuss problems but simply told each other the answers or worked individually. All questions were initiated by the teacher. Students did not have an opportunity and were not encouraged to ask questions of each other.
C. Target-centered questions **Score**	The use of strategic or target-centered questions for formative assessment was intentional and clearly planned; all student responses were considered and used to adjust the pace and focus of the lesson.	The use of some target-centered questions for formative assessment was intentional, with some evidence of planning for them; many, but not all, student responses were considered and/or utilized to alter the pace or focus of the lesson.	The use of some target-centered questions for formative assessment was intentional, but with little evidence of intentional planning for them; many, but not all, student responses were considered and/or utilized to alter the pace or focus of the lesson.	The use of target-centered questions for formative assessment was limited, with no evidence of planning for them; some student responses were considered and/or utilized to alter the pace or focus of the lesson.	The intentional use of target-centered questions for formative assessment was not evident; responses by students did not alter the pace or focus of the lesson.
D. Feedback to Responses **Score**	Students always had ample time to consider the question before responding; appropriate feedback (e.g., clear, specific, and descriptive) was consistently given to all students; feedback always encouraged student involvement in the discussion/task.	Students usually had ample time to consider the question before responding; appropriate feedback (e.g., clear, specific, and descriptive) was given to most of the students; feedback usually encouraged student involvement in the discussion/task.	Students generally had time to consider the question before responding; however, response time varied and/or was not consistent; appropriate feedback (e.g., clear, specific, and descriptive) was given to many of the students; feedback did not always encourage student involvement.	Students had minimal time to consider the question before responding; appropriate feedback (e.g., clear, specific, and descriptive) was given occasionally or seemed to be given only to a few students; limited encouragement for students to become involved.	Students had no time to consider the question before responding; appropriate feedback (e.g., clear, specific, and descriptive) was not given or given only to a few students; no encouragement for student involvement.
Section III Comments					

IV. Classroom Atmosphere					
A. Student Involvement **Score**	All of the students demonstrated interest and were engaged in the instructional activity.	Most of the students demonstrated interest and were engaged in the instructional activity.	Approximately equal numbers of students were interested/engaged and not interested/engaged in the instructional activity.	Only a few of the students were interested/engaged in the instructional activity.	None of the students were interested/ engaged in the instructional activity.
B. Classroom Management **Score**	The classroom was well managed and totally orderly; there were no student disruptions which caused a loss of instructional time or impaired the learning environment.	The classroom was well managed and orderly; some minor student incidents which did not require any corrective action and did not cause any loss of instructional time.	The classroom generally well managed and orderly; one or a few minor student disruptions occurred which required corrective or disciplinary action causing a minimal loss of instructional time.	The classroom was poorly managed and/or disorderly with frequent student disruptions that required corrective or disciplinary action and caused a significant loss of instructional time.	The classroom was disorderly with constant student disruptions that caused a major loss of instructional time and seriously impaired the learning environment.
C. Classroom Culture **Score**	The teacher has established a classroom culture in which all, or nearly all, of the students take initiative in discussions and activities; all students demonstrated respect for other students; all, or nearly all, demonstrated enthusiasm, confidence, persistence and accuracy while solving problems.	The teacher has established a classroom culture in which most of the students take initiative in discussions and activities; most students demonstrated a respect for other students; most students demonstrated enthusiasm, confidence, persistence and accuracy while solving problems.	The teacher has established a classroom culture in which many (majority) of the students take initiative in discussions and activities; most students demonstrated respect for other students; many students demonstrated some attitudes such as enthusiasm, confidence, persistence and accuracy while solving problems.	The teacher has established a classroom culture in which only a few students take initiative in discussions and activities; the majority demonstrated a respect for other students; a few students demonstrated enthusiasm, confidence, persistence and/or accuracy while solving problems.	The teacher has established a classroom culture in which students did not feel comfortable taking the initiative in discussions and activities; few, if any, demonstrated a respect for other students; no, or only a very few, students, demonstrated enthusiasm, confidence, persistence or accuracy in their work.
Section IV Comments					
V. Analysis of Instruction Leading to the development of Higher Order Mathematics Skills					
A. Amount/ Level of Student Problem Solving	Students were engaged in problem-solving activities that utilized higher level thinking skills. Students solved or investigated	Students were engaged in problem solving activities that utilized higher level thinking skills. Students solved teacher or student-	Students were engaged in problem solving activities that used some higher level skills; however, the focus of lesson was on	Students were involved in only low level problem solving or were not involved in any type of problem solving/	Students were not involved in any type of problem solving activity; there was no evidence that the content/concept

Score	teacher or student-initiated problems using effective and innovative strategies. The problems required students to analyze data, generalize to make conjectures, justify solutions and/or connect math with real world situations.	initiated problems using effective strategies. The problems allowed students to analyze data, generalize to make conjectures and/or connect mathematics and real world situations.	basic skills. Students solved typical problems that required only a definitive procedure to be correct; the problems allowed students to apply theorems, collect/analyze data or justify solutions.	investigative activity; the lesson objective did not require such activity but would have been enhanced by it.	was learned using the strategies employed in this lesson.
B. Math skills being developed **Score**	Students were developing/ utilizing higher level mathematics skills during the entire class period; interpretive discussions involved all students at a high-level.	Students were developing/ utilizing higher level mathematics skills during most of the class period; interpretive discussions involved most of the students at a high-level.	Students were observed utilizing higher level skills, however, the focus of the instruction was on lower level skills during most of the period; interpretive discussion was minimal or engaged only part of the students.	Students were not observed utilizing higher level skills; the entire focus of the lesson was on lower level mathematics skills with minimal or no interpretive discussion.	Students were not engaged in any activity in which either higher or lower level mathematics skills were developed; no interpretive discussion was observed.
Section V Comments Students were engaged in an authentic problem solving experience and were required to analyze and interpret real-world data. They were actively involved in higher-cognitive discussion both in groups and whole class.					
VI. The Teacher Demonstrates Applied Content Knowledge					
A. Communication **Score**	Consistently used accurate and effective communication; vocabulary was clear, correct and appropriate.		Generally used accurate and effective communication; occasionally used of inappropriate vocabulary; exhibited some minor errors that did not interfere with conceptual development		Consistently used inaccurate, misleading and/or ineffective communication and/or inappropriate vocabulary.
B. Connects Content to Life Experiences **Score**	Consistently connected content, procedures, and activities with relevant life experiences or current events.		Connected some content, procedures, activities with relevant life experiences or current events.		Rarely/ never connected content, procedures or Activities with relevant life experiences or current events.

C. Instructional Strategies Appropriate for Content and Contribute to Student Learning **Score**	Used instructional strategies that were clearly appropriate for the content/processes of the lesson; evident that student learning occurred as a result of the strategies employed.		Used instructional strategies that were generally appropriate for the content/processes of the lesson; however, not clear if the student learning was a result of the strategies employed.		Used instructional strategies that were questionable or inappropriate for the content/processes of the lesson; no indication that student learning occurred.
D. Guides Students to Understand Lesson Content from Various Perspectives **Score**	Provided multiple opportunities for students to consider content from different perspectives or contexts.		Provided one or a few Opportunities for students to consider content from different perspectives or contexts.		Rarely/never provided opportunities for students to consider content from different perspectives or contexts.
Section VI Comments					
VII. The Teacher Creates and Maintains a Positive Learning Climate					
A. Communicates High Expectations **Score**	Presented significant and challenging objectives and consistently communicated confidence in students' ability to achieve.		Presented challenging objectives; and at times communicated confidence in students' ability to achieve.		Presented minimal or no objectives for students; rarely or never communicated confidence in students' ability to achieve.
B. Establishes a Positive Learning Environment **Score**	Clear conduct standards have been established and are being met; awareness of student behavior; responded appropriately/ respectfully to student misbehavior.		Conduct standards have been established; there was some inconsistency in monitoring and response to student misbehavior.		No established conduct standards or expectations; minimal or no monitoring; inappropriate responses to student misbehavior.

C. Values and Supports Student Diversity (including gender ethnicity, S.E.S., academic and physical abilities) **Score**	Recognized and consistently responded to the diversity in the class; consistently used or attempted to use strategies to address the needs of all students.		Recognized but inconsistently responded to the diversity in the class; used or attempted to use some different strategies to address the needs of particular students.		Provided little or no recognition or response to student diversity and individual needs; used the same approach for all students.
D. Fosters Mutual Respect Between Teacher and Students and Among Students **Score**	Always treated all students with respect; encouraged and clearly expected students to treat each other with respect.		Generally treated students with respect; provided some encouragement of students to treat each other with respect.		No evidence of the teacher's respect or concern for students was observed; provided little or no encouragement of students to treat each other with respect.
E. Provides a Safe Environment for Learning **Score**	Classroom environment was emotionally and physically safe for students at all times.		Classroom environment was emotionally and physically safe for students most of the time.		Classroom environment was not emotionally and/or physically safe for students.
Section VII Comments					
VIII. The Teacher Implements and Manages Instruction Leading to Positive Student Outcomes					
A. Implements Instruction Based on Student Needs and Assessment Data	Instruction addressed all individual student needs; always used or attempted to use a variety of appropriate instructional strategies to meet individual	Instruction addressed most individual student needs; used different instructional strategies as needed to meet needs of most students; sometimes adapted instruction to meet	Instruction addressed many individual student needs; used more than one instructional strategy as needed; occasionally adapted instruction to meet changing or	Instruction addressed some individual student needs; attempted to use more than one instructional strategy; seldom adapted instruction to meet	Instruction did not address individual student needs; one strategy was used for all students; no attempt to adapt lesson to meet changing or

Score	student needs; adapted instruction to changing or unanticipated circumstances	changing or unanticipated circumstances	unanticipated circumstances	changing or unanticipated circumstances	unanticipated circumstances.
B. Uses Time, Space , and Materials Effectively **Score**	Always used efficient procedures for non-instructional tasks; no loss of learning time was observed; classroom space and materials were always used effectively to facilitate student learning.	Used efficient procedures for non-instructional tasks most of the time; minimal loss of learning time was observed; classroom space and materials were used effectively to facilitate student learning.	Generally used efficient procedures for non-instructional tasks with some loss of learning time; Classroom space and materials were used effectively most of the time.	Used both efficient and inefficient procedures for non-instructional tasks resulting in significant loss of learning time; classroom space and/or materials were used effectively to facilitate student learning some of the time	Used inefficient procedures for non-instructional tasks resulting in major loss of learning time; classroom space and materials were not used effectively to facilitate student learning.
C. Implements and Manages Instruction to Facilitate Higher Order Thinking **Score**	Instruction encouraged higher order thinking of all students; included significant amount of independent and/or group processing and reflection time.	Instruction encouraged higher order thinking of most students; included some independent or group processing and reflection time.	Instruction encouraged higher order thinking by some students; included minimal independent or group processing and reflection time.	Instruction encouraged higher order thinking by only a few students; little, if any, independent/ group processing or reflection time was provided.	Instruction was minimal and ineffective; did not encourage higher order thinking by any students; did not include any independent/ group processing or reflection time
Section VIII Comments					
IX. The Teacher Assesses and Communicates Learning Results					
A. Uses Assessments Aligned with Learning Objectives **Score**	Formative assessment strategies were fully aligned with learning objectives; assessment results were obviously used to adjust instructional practice in a timely manner.	Formative assessment strategies were aligned with learning objectives; appeared to be used to adjust instruction.	Formative assessment strategies were generally aligned with learning objectives; not clear if or how assessment results were used to adjust instructional practice.	Formative assessment strategies not clearly aligned with learning objectives; appeared to be done without intention or done for compliance only.	No assessment strategies were used even though formative assessment was needed to determine the level of student learning

B. Uses a Variety of Formative and Summative Assessments to Measure Learning **Score**	Used various formative and/or summative assessment strategies that provided all students several opportunities to demonstrate learning.	Used various formative and/or summative assessment strategies that provided most students some opportunities to demonstrate learning.	Used some formative and/or summative assessment strategies that provided many students (at least the majority) opportunities to demonstrate learning.	Limited use of formative and/or summative assessment strategies that provided opportunities for some students to demonstrate learning.	No assessment strategies were used even though formative assessment was needed to determine the level of student learning.
C. Adapts Assessments to Accommodate Diverse Learning Needs/ Situations **Score**	Formative and/or summative assessment strategies were obviously adapted to accommodate student diversity and diverse learning needs.	Formative and/or summative assessment strategies were appeared to be adapted to accommodate student diversity and diverse learning needs.	Some attempts were made to adapt formative and/or summative assessment strategies to meet diverse needs; however, these were not successful for all students.	A limited attempt was made to adapt assessment strategies to accommodate student diversity or to meet diverse student needs.	No assessment strategies were used even though formative assessment was needed to determine the level of student learning.
Section IX Comments					
X. Overall Classroom Observation Rating					
Overall rating of quality of instruction **Score**	Instruction was of high quality and effective for all students; there was evidence that instruction was based on clearly defined objectives that were fully aligned with standards; all students were engaged in activities requiring higher level thinking skills.	Instruction was of high quality and effective for most students; there was evidence that instruction was based on clearly defined objectives that were aligned with standards; most students were engaged in activities that required higher level thinking skills.	Instruction was of good quality and effective for many students; instruction was based on student objectives somewhat aligned with standards; some students had an opportunity for higher level thinking skills development.	Instruction was of mediocre quality and effective for only a small portion of the students; little evidence that instruction was based on student objectives; instruction had minimal impact on student learning.	Instruction was of poor quality and was not effective for any students; no evidence that instruction was based on student objectives; learning was not based on instruction provided.
Section X Comments					

Physical Setting – This information is included for contextual purposes only. These data are not included in the rubric ratings.			
A. Classroom facilitates student learning **Score**	Flexible student furnishings can accommodate any type of mathematics activity to provide for maximum student and/or teacher interactions.	Student furnishings are somewhat flexible and can accommodate interaction for most types of mathematics instructional activities.	Student furnishings are not flexible and in many cases limit the interactions needed for quality mathematics instruction.
B. Classroom facility **Score**	Classroom is large with sufficient storage for supplies; materials are well organized for ease of access; classroom furnishings are appropriate for problem solving and/or hands-on activities and materials are available for all students.	Classroom is generally adequate in size with some storage; materials generally well organized for ease of access; classroom furnishings are generally appropriate for problem-solving and/or hands-on activities but may not accommodate all students at the same time.	Classroom is inadequate in size with little or no storage; little, if any, organization of materials allows ease of access; classroom furnishings are not conducive to problem-solving instruction or can only accommodate small numbers.
C. Classroom environment **Score**	Mathematics instructional resources are abundant and easily obtained, many mathematics displays promote learning, and student work is prominently displayed.	Mathematics instructional resources may be available but not in sufficient quantities, e.g., graphing calculators; some mathematics posters or displays that promote learning were observed, and limited amount of student work is displayed.	Mathematics instructional resources are absent or extremely limited; no evidence of mathematics displays and no student work was posted.
Physical Setting Comments			

November, 2013

Appendix C. A pilot test of the Leadership by Design scoring rubric for assessment of instructional quality

Linda Cavalluzzo, Stephen Henderson, and Christine Mokher
January 2010

Overview

This is a report on a teacher observation pilot conducted for the study *"The Contribution to Teacher Effectiveness of National Board Certification"*, which will examine the impact of on teaching effectiveness of going through the National Board Certification (NBC) application process. One aspect of the evaluation involves a comparison of classroom observations from a sample of teacher NBC candidates with similar teachers not pursuing this certification. The goal of this part of the study is to chart the observed use of effective instructional practices as teachers move through the NBC process as compared to non-NBC applicant teachers with similar characteristics in similar classroom settings. Changes in instructional quality will be examined for science teachers in 34 schools in Kentucky (17) and Chicago (17) over a three-year period. Growth in instructional quality for NB-involved teachers will be compared to teachers who are not involved with the NB process to draw conclusions about the gains in instructional quality made by science teachers as a result of participation in the certification process.

This study design requires the use of a comprehensive observation instrument to document what is observed, a tool for assigning numeric scores to the instructional practices observed, and consistent and reliable data collection and scoring procedures to maintain the internal validity of these data. The Leadership by Design (LBD) Science Classroom Observation Instrument, modified to ensure consistency with the NB science standards, has been selected for use in the study. This instrument has been widely used in Kentucky and elsewhere; classroom observation data have been collected using the LBD instrument for over 3,000 teachers in more than 250 elementary, middle, and high schools in 7 different states. Projects utilizing the LBD include work funded by the U.S. Department of Education and the National Science Foundation. The LBD has also been adopted by the National Science Teachers Association as a program improvement tool to help assess and improve the quality of instruction in middle school and high school classrooms.

In contrast to the extensively used LBD, the instrument for assigning numeric scores to the observation data —the LBD Classroom Observation Rubric—was newly developed for this study. Thus, we have conducted pilot observations for a small sample of science teachers to identify any problems transferring the observation data to the rubric and to ensure that the

scoring data are internally consistent. We are using these data to identify and address any issues with the rubric itself or with the procedures of translating the LBD instrument data into our scores on the rubric prior to conducting the observations for the study. In the actual study, the scoring of instruction will be based on classroom observations and supporting information obtained from the teacher debrief interviews and a review of lesson plans and sample assessments.

For this pilot study, the developer of LBD, Dr. Henderson, trained five observers in use of the LBD and scoring rubric. As described in more detail in the appendix [Table 11] to this report, the five observers are all experienced science educators who also have used the LBD instrument for previous studies. The observers did not collect additional materials or debrief the teachers following the classroom observation, as they will be expected to do to increase the reliability of the results in the actual study.

Completed LBD observation instruments and scoring rubrics were collected by Dr. Henderson from the classroom observers following their classroom visits. Copies of the completed data collection instruments were provided to CNA for independent analysis.

Summary of Major Findings

Overall, no major concerns were identified with the use of the rubric in the pilot observations. Using the LBD and scoring rubric,

- *Observers were able to distinguish the level of instructional quality among science classrooms.*
    - 56 percent of the individual items rated on the scoring rubric (N=21) had ratings that covered the entire range of possible scores from 1 to 5.
    - All individual items had a range of at least two points.
    - Among the 9 subscales, the minimum scores were between 1.0 and 2.7, while the maximum scores were between 4.7 and 5.0.

- *Missing data were minimal.*
    - For 7 out of 11 of the scales and subscales, none of the 9 observed classrooms had any N/A or missing ratings.
    - For 3 out of 11 of the scales and subscales, there was 1 observation with a N/A or missing rating for 1 item.
    - For the remaining subscale, "IX. Assesses Learning", 4 of the items were rated as "N/A" or were missing a mark on the LBD for one or more items. However, during the pilot, observed teachers were not asked to provide a sample assessment for review by the observer. In the actual study teachers will be asked in advance to have this information on hand for the observer.

- *Overall ratings were consistent with ratings on subscales.*
    - The overall rating for quality of instruction is not expected to be the average of the subscales. Nevertheless, we would expect that teachers who receive high overall ratings to tend to have high ratings on each of the subscales. We found no anomalies in the ratings.

        - The average rating for each of the subscales between teachers with low overall instructional ratings (score of 2 or 3) and high overall instructional ratings (score of 4 or 5) were 0.7 to 2.9 points higher for teachers with high overall ratings.

- *Scores exhibited high face validity*
    - The pilot sample consisted of classroom observations for nine science teachers, two NBCTs and seven others. Observers were not aware of which teachers were the NBCTs. Because NBCTs have been certified for the quality of their professional practices, we would expect them to score well on the LBD and higher, on average, in comparison to teachers who have not gone through the NB process. We found that,
        - Both of the National Board teachers had an overall instructional rating of 5, the highest possible rating, compared to a mean of 3.3 for the Non-National Board teachers.
        - The average subscale ratings for the National Board teachers were also higher than the Non-National Board teachers' average rating for each of the 9 subscales.

A second training session for the observers will be conducted before the actual data collection to ensure that observers have been refreshed on how to score the rubric and to address a few minor issues that were uncovered during the pilot observations, which will be discussed in more detail in this report.

Below we provide more detail on major findings from the pilot study and describe the observation process, examine variation in scores, document the extent of missing data and items marked N/A, provide context for understanding the overall ratings, examine the internal consistency of the ratings, assess the face validity of the results, demonstrate how sample results may be displayed in the final report, and discuss the conclusions and implications for the study.

Observation Process

The observation team for the study consists of seven experienced science educators who have been trained in the use of the LBD instrument and have conducted observations in actual classroom environments. Though our observers are experienced and well-qualified, we provided a full-day training session in October on the NBC-LBD Classroom Observation Instrument and the LBD Classroom Observation Rubric. Prior to the start of study observations, further training will be provided to address and correct the issues identified in this pilot study.

For these pilot observations, classroom observations were collected from a nonrandom sample of nine (9) middle school and high school science teachers in Blount County (Maryville), Tennessee and Fayette County (Lexington), Kentucky. Two of these teachers had National Board certification. The pilot observations were conducted by five of our

trained observers who all have previous experience teaching science and conducting classroom observations, as described in the appendix [Table 11].

During the classroom observation, the observer filled out the LBD instrument, marking items as they were observed. Observers were instructed to mark a response for every item. Following each classroom observation, the observer reflected on the observation and, using the completed LBD instrument, filled out the LBD Classroom Observation Rubric. Note that the LBD acts as a memory device for the observer when filling out the scoring rubric; the data collected from the LBD are not used directly in the study. For actual study observations, the observer will also obtain planning materials and assessments from the teacher, and conduct a short debrief following the observation. These materials and the discussion with the teacher will enable the observer to better understand what was observed, facilitating more accurate completion of the rubric.

Each item on the rubric is scored on an integer scale of 1-5, with 5 being the highest rating and 1 the lowest. If an item was not observed, it is marked as "Not Applicable" and is not assigned a numeric score. The rubric consists of 9 instruction-related subscales which are based on the average rating of 3 to 5 specific items aligned with the LBD instrument. The rubric also has an overall quality of instruction rating, and a subscale for the physical setting. The physical setting rating is collected to provide baseline contextual information and is not used to evaluate the teacher or quality of instruction.

Variation in Scores

Sufficient variation in scores is needed to distinguish differences in teachers' instructional quality. We examined the distribution of scores for each item and subscale on the rubric. Fifty-six percent of the individual items (N=21) had ratings that covered the entire range of possible scores from 1 to 5. All individual items had a range of at least two points. Among the 9 subscales, the minimum scores were between 1.0 and 2.7, while the maximum scores were between 4.7 and 5.0 (see Table 9). The range of ratings for all subscales was between 2.3 and 4.0, out of a possible 5-point scale. The distribution of scores was similar for the Physical Setting and Overall Rating. These findings indicate variation is present in all of the ratings.

Missing Data and Items Marked "N/A"

Missing ratings or items marked as "N/A" were excluded from the averages that were calculated for each subscale. If too many of these items are excluded from a subscale, then the corresponding rating may not be a reliable indicator of the construct it is designed to measure. A teacher's average rating on a subscale may also be disproportionately influenced by the score on a single item if data are missing for other items in the scale. We checked for patterns in missing and "N/A" ratings by examining which items were most commonly missing and whether any individual observers reported an unusually high number of missing or "N/A" ratings. Any issues identified may indicate a need to revise specific items or provide additional training to the observers about how to score them.

Table 9: Subscale, physical setting, and overall rating statistics (total number of items, minimum, maximum, range, and average).

	Total # Items	Minimum	Maximum	Range	Average
I. Lesson Overview	5	2.2	5.0	2.8	4.0
II. Instructional Overview	4	1.0	5.0	4.0	3.6
III. Questioning	4	2.0	5.0	3.0	3.7
IV. Classroom Atmosphere	3	2.7	5.0	2.3	4.2
V. Higher Order Skills	3	2.0	4.7	2.7	3.3
VI. Content Knowledge	4	2.5	5.0	2.5	3.5
VII. Positive Climate	5	2.5	5.0	2.5	4.1
VIII. Implements Instruction	4	2.5	5.0	2.5	3.7
IX. Assesses Learning	3	1.5	5.0	3.5	3.4
Physical Setting	3	1.7	5.0	3.3	4.1
Overall Rating	1	2.0	5.0	3.0	3.7

Table 10 shows the number of teachers who had N/A or missing ratings for 0, 1, 2, or 3 items for each of the scales and subscales. For 7 out of 11 of the scales and subscales, none of the 9 teachers had any N/A or missing ratings. For 3 out of 11 of the scales and subscales, there was 1 teacher with a N/A or missing rating for 1 item. The remaining subscale, "IX. Assesses Learning", was more problematic, with 4 of the teachers marked as "N/A" or missing for one or more items.

Table 10: Total number of items, and number of teachers with N/A or missing ratings for 0, 1, 2, or 3 items: by scale or subscale

	Total # Items	# Teachers with N/A or missing ratings for:			
		0 items	1 item	2 items	3 items
I. Lesson Overview	5	9	0	0	0
II. Instructional Overview	4	9	0	0	0
III. Questioning	4	9	0	0	0
IV. Classroom Atmosphere	3	9	0	0	0
V. Higher Order Skills	3	8	0	1	0
VI. Content Knowledge	4	8	1	0	0
VII. Positive Climate	5	7	2	0	0
VIII. Implements Instruction	4	9	0	0	0
IX. Assesses Learning	3	5	1	1	2
Physical Setting	3	8	1	0	0
Overall Rating	1	9	0	0	0

NOTE: A total of 9 teachers were observed.

However, during the pilot observations the teachers were not asked to provide a sample assessment in advance, so the observers may have been unable to assign a rating for these

items if the teacher did not have an assessment available to review. Before the actual observations are conducted the teachers will receive a letter with instructions regarding materials they should have available, so we do not anticipate the same problem with N/A and missing ratings for this subscale.

Context for Understanding Overall Ratings

After rating each of the items on the rubric, observers were asked to assign an overall rating of quality of instruction. This rating is designed to take into account the observer's overall impression including the effectiveness of instruction, alignment with objectives and standards, student engagement, and development of higher order thinking skills. We examined the observer's written comments and responses on the LBD instrument to provide some context for understanding what was scored. Below are two examples of how the classroom observations corresponded to the overall instructional ratings.

Teacher 1 taught a lesson entitled "What is friction?" to a grade 7 science class. The objective of the lesson was to describe how the mass of an object can affect the outcome of collisions. The students worked in small groups on a lab assignment to study collisions using marbles. However, mass was never measured, only judged by the size of a marble. The student investigations focused on basic skills such as "observing" and "inferring" instead of higher level skills like "formulating hypotheses" and "interpreting data." Despite these limitations, nearly all of the students were engaged, the learning objectives were clearly communicated using multiple means, the teacher communicated effectively with the students, and a formative assessment was observed during a closure discussion with the whole class. The teacher received an overall instructional rating of 3.

Teacher 3 taught a lesson on enzymes to a high school biology class. Students worked in small groups to investigate how variables affect enzyme activity by designing and performing their own experiments. The emphasis in these investigations was on higher-level skills such as "evaluating data" and "interpretive discussion." All students were encouraged to ask questions, and the questions stimulated higher level and divergent thinking. The observer described the classroom culture as "enthusiasm for learning" and "curiosity." The teacher clearly communicated the learning objectives using multiple means and used formative assessment that was fully aligned with these learning objectives. The teacher received an overall instructional rating of 5.

The report will also describe differences in the types of activities observed in the classrooms of teachers with high ratings and low ratings. The LBD instrument asks observers to identify both the instructional strategies used by the teacher and the activities performed by students during the class.

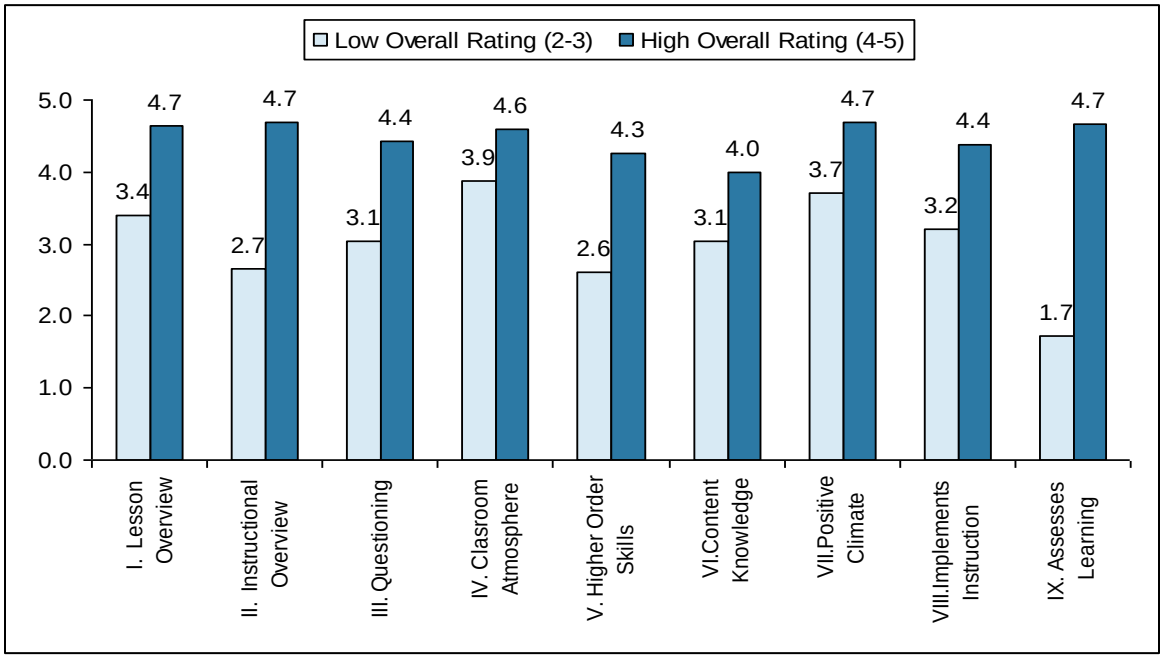
Internal Consistency

The overall rating for quality of instruction is not expected to be the average of the subscales. For example, suppose a teacher has clear objectives, assigns activities that promote higher level skills, asks challenging questions, and demonstrates strong content knowledge,

but none of the students are engaged or following the lesson. The teacher would likely receive high ratings for many of the subscales except "classroom atmosphere", so the average of the subscale ratings would be relatively high. However, if the students do not appear to be learning much from the lesson, the observer may perceive a lower overall quality of instruction.

Even though there is not an exact match between the overall instructional rating and the average of the subscales, we would expect that teachers who receive high overall ratings would tend to have high ratings on each of the subscales. In order to examine the internal consistency of the ratings, we compared the average rating for each of the subscales between teachers with low overall instructional ratings (score of 2 or 3) and high overall instructional ratings (score of 4 or 5). The subscales for teachers with high overall ratings were 0.7 to 2.9 points higher compared to teachers with low overall ratings (see Figure 10).

Figure 10: Average rating on each subscale for teachers with low and high overall instructional ratings.

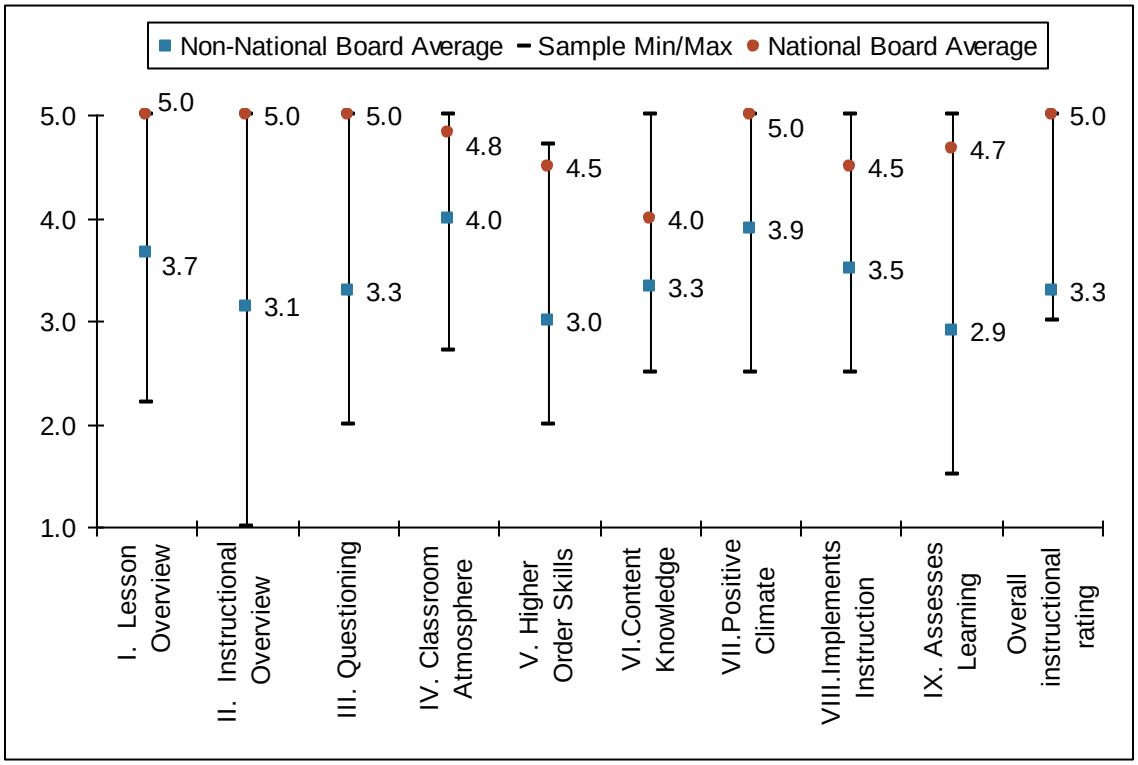


Face Validity

Face validity is conducted by examining outcomes to consider whether a measure appears to assess what it is designed to assess. In the early stages of selecting an instrument for the observation, a crosswalk was created to show that many of the same standards used in National Board certification are captured on the LBD instrument. Thus we would expect that teachers with National Board certification should score highly on the ratings from the classroom observations in this study.

Two of the nine teachers observed in the pilot observations were National Board certified, although the observers did not know of the teachers' certification status until after the observations were conducted. Figure 11 shows how the average of the National Board teachers' ratings compared to the mean for Non-National Board teachers, as well as the minimum and maximum ratings for the sample. Both of the National Board teachers had an overall instructional rating of 5 compared to a mean of 3.3 for the Non-National Board teachers. The average subscale ratings for the National Board teachers were also higher than the Non-National Board teachers' average rating for all 9 of the subscales.

Figure 11: Comparison of ratings for the Non-National Board average, sample minimum/maximum, and National Board average, by subscale.



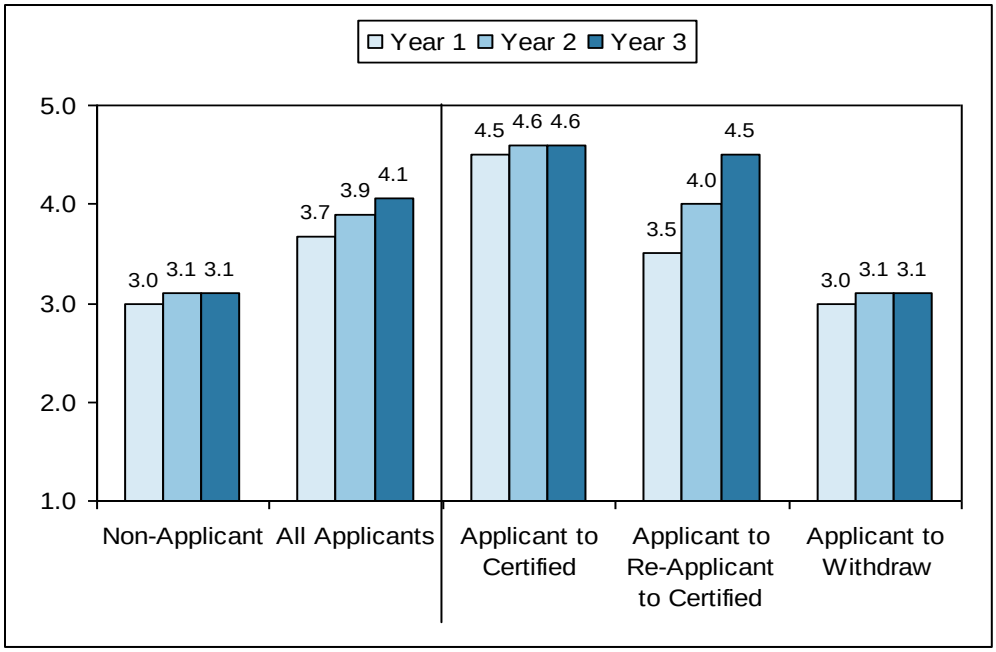
Sample Results

The ratings from the pilot observations represent observations taken at a single point in time, whereas the study will track teachers over time and will include an average of two observations per teacher in each time period. The final report will show how the ratings changed for teachers with different types of National Board participation. Tests of statistical significance will be conducted to determine if there are differences in the change over time between teachers with no involvement and teachers with various level of involvement in the certification process.

Figure 12 provides an example of how the information may be displayed graphically for the overall rating of instruction. In this hypothetical case, the teachers with no involvement in

the National Board process begin with an average rating of 3.0 in year 1, and show little improvement over time with subsequent scores within 0.1 points. Across all National Board applicants, there is a 0.4 point increase over time from 3.7 to 4.1. However, when the results are disaggregated among different stages of applicants, the increase in scores is largely attributed to a single group. The teachers who changed from "new applicant to re-applicant to certified" demonstrated the greatest growth, with average ratings of 3.5 in year 1, 4.0 in year 2, and 4.5 in year 3, for a total change of 1.0 point over three years. The teachers whose status changed directly from "applicant to certified" were good teachers in the beginning of the process and did not change much over time, with average scores between 4.5 and 4.6 in all years. The teachers whose status changed from "applicant to withdraw" had similar ratings to the non-applicant teachers, with a rating of 3.0 in year 1 and 3.1 in years 2 and 3.

Figure 12: Sample figure for overall instructional ratings in years 1, 2, and 3; by National Board participation status (Hypothetical data).



Conclusion & Implications

Overall, there were no major issues with the use of the rubric in the pilot observations. The ratings revealed variation in scores across teachers, there were no systematic patterns of missing data, the observer comments reflected the corresponding ratings, there was internal consistency between the overall instructional ratings and the subscale ratings, and face validity was established among observed outcomes.

The research team has identified several changes that should be made before the observations for the study are collected. The changes include the following:

- Requiring observers to write comments corresponding to the overall rating to provide context for understanding why the rating was selected.
- Reminding the observers to check the consistency between the LBD instrument and the scoring rubric. One observer assigned an overall rating of 4 on the instrument and 5 on the rubric for the same teacher.
- Emphasizing the importance of collecting sample assessments from the teacher so that ratings can be assigned to the section on assessment.
- Changing the instructions on the "Instructional Overview" section of the instrument from "Mark only one" to "Mark all that apply" for the instructional resources used. Observers may still be asked to distinguish which activities or strategies were primarily used, but selecting all that were observed will provide a better understanding of what occurred in the lesson
- Reviewing with the observers how to score sections on the LBD that are not as closely aligned with the rubric. For a few of the individual items on the subscales (e.g. 1c "Content Delivery"), reviewers checked similar boxes on the LBD but there was variation in the corresponding ratings on the rubric. Observers will spend more time discussing these items on the rubric that do not directly match the LBD so there is a shared understanding about how these items should be rated. Observers will also be asked to provide written comments to explain any cases where the marks on the LBD appear favorable but the rubric rating is low, and vice versa.

A second training session for the observers will be held prior to data collection for the study. At that time, results of the pilot observations will be debriefed and the items listed above will be reviewed. Observers also will be asked if they encountered any problems transferring the observation data to the rubric and whether any additional data should be collected on-site to better reflect the classroom observation experience.

(Appendix) Professional Background of Classroom Observers

The table below provides a summary of the professional experience of the five observers used in this pilot test of the data collection instrument. For the actual data collection, a total of seven observers are planned. Two NBCTs are being sought to fill out the observation team.

Table 11: Science education experience, LBD involvement, and related experience of the 5 observers who participated in the pilot.

Science Education Experience	LBD Involvement	Related Experience
14 years science education experience -high school physics teacher in Tennessee, Pennsylvania and West Virginia; School/ District Consultant for Math/Science Partnership Projects	Worked with school districts in Tennessee training principals in collection and analysis of classroom observation data using LBD system.	Adjunct Faculty Member, TN postsecondary school– Taught physics and physical science courses for future teachers; Coordinator of US DOE funded curriculum development project.
14 years of experience as high school science teacher, Kentucky; Science Content Specialist, KY school district.	Utilized the LBD program for classroom observations in KY; utilized LBD data to analyze program improvement efforts	Adjunct Faculty Member, Kentucky postsecondary school. Taught high school science education methods courses
30 years science education experience - biology/physical science teacher in Missouri and Tennessee; University professor of science education; Director of math/science partnership projects; Owner/Executive Director of science/mathematics program improvement consulting firm	Utilized the LBD program as part of federal program development work; trained as a Program Improvement Profile observer using the LBD program; worked with school districts in Tennessee training principals in collection and analysis of classroom observation data using LBD system.	Education Partnerships Team/ Program Leader for large U.S. corporation; Director of federal science resource collaborative at Univ. of Tennessee; Assistant Professor of Science Education, TN postsecondary school.
28 Years as high school and middle school science teacher in, large KY school district	Classroom observer using the LBD program for the past 12 years; Certified Reviewer for the NSTA Science Program Improvement Review which utilizes the LBD instrument	Regional manager of Partnership Reform Initiatives in Science and Math – NSF funded project; Consultant for a Kentucky High School Math Science Partnership technology education project; Science Education Consultant with a regional cooperative
30 years as an elementary and middle grades science teacher in two KY school districts	Classroom observer using the LBD program for the past 12 years; Certified Reviewer for the NSTA Science Program Improvement Review which utilizes the LBD instrument	Regional manager of Partnership Reform Initiatives in Science and Math – NSF funded project; Science Education Consultant for a federal initiative focusing on school improvement initiatives; National Presidential Award for Excellence in Science Teaching

Appendix D. Construction of the student analytic file

The statistical analysis of administrative data examines the impact of the National Board certification process on teachers' effectiveness in increasing their students' test scores. Student-level data files were collected for SYs 2007/08, 2008/09, 2009/10, and 2010/11 from KDE; and for SYs 2008/09, 2009/10, 2010/11, and 2011/12 from CPS. These student-level data are provided in several different files, include school enrollment records, student demographic records, student course transcripts, and student test scores (EXPLORE, PLAN, and ACT). The course transcript file includes records for all of the courses that each student took in the corresponding year, with the teacher of record for each of those courses.

NBPTS provided a teacher-level file with records on new NBC applicants in the 2001–2002 application cycle through the 2011–2012 application cycle. The variables in this file include teacher names and email addresses, school and district names, cohort, certificate type, cycle date, application date, and certification status. We used the teacher names, school names, and email addresses to match these records to the teachers in the student-level course data from KDE and CPS.

We combined the data described above to create one longitudinal file for Kentucky and one for Chicago Public Schools. Students who are missing records on the pretest and/or posttest variables were dropped from the sample. Each file in Kentucky has multiple records per student in each subject that correspond to records for the semesters from the administration of the pretest through the administration of the posttest. The CPS file includes up to two records per student, one each for the PLAN and ACT analyses. Records corresponding to the PLAN analysis include grade 9 classroom teachers in the core subject, as well as PLAN and EXPLORE test scores. The records corresponding to the ACT analysis include both grade 10 and grade 11 classroom teachers in the core subject areas, as well as ACT

and PLAN test scores. This allows us to attribute gains in student achievement to all of the teachers who taught a student in the time from the pretest to the posttest. Standardized state or district course codes were used to categorize courses into three subject areas: English, math, and science.

Handling students with missing teachers

One issue we encountered when constructing the data file is that not all students take a course in the same subject area for all semesters (for all years) between the pretest and the posttest. For these cases, we created a new record for the missing periods and assigned a "missing" teacher ID so that these students could be included in the analyses.

When we examined the records missing teachers more closely, we found that one of the reasons this was occurring in Kentucky was because some schools are on a block schedule. Block courses meet more frequently during the week or have class periods with a longer duration than traditional courses in order to allow students to receive a full year of credit in a single semester.

Some 37 percent of students in the Kentucky sample had a block course in at least one of the semesters between the pretest and the posttest. For these cases, we created a dummy "block" variable in the semester that the block course was taken to indicate that the teacher was teaching a block course. If no course was taken in the other semester of the same school year, we created a new record for this semester with a missing "block" teacher ID.

There were also some students who took block courses in the same subject in both the fall and spring semesters in a single academic year. These students experienced twice as much instructional time as students in a traditional yearlong course. We created a separate dummy variable ("double block") to indicate that these students completed two block courses in the same year. Only 2 percent of students in the Kentucky sample were in this category.

Handling students with multiple teachers

Another issue that we encountered when creating the data file was that some students have more than one teacher in the same subject area in a single semester. One reason this can occur is because stu-

dents complete both a core course and an elective course in the same subject (e.g., a core English course and a creative writing course). For Kentucky, we reviewed the descriptions of the state course codes and categorized courses as core if they counted toward the state graduation requirements in the subject, or elective otherwise. For CPS, we used the descriptions from the CPS graduation requirements. We created a dummy variable to indicate whether students had an elective course in the same subject area as the core course, and then dropped the records for the elective courses.

There are other reasons why students might have multiple teachers. Some students may switch classes or schools in the middle of the semester. Other students may enroll in more than one core course in the same subject area in a single semester. In all of these cases, it is difficult to identify to which teacher to attribute changes in student outcomes. We created a new record for the semester in which this occurred, and assigned a missing "multiple" teacher ID variable that indicates the student had multiple teachers during the semester.

Finally, in CPS we had some teachers who match to at most five students in the analysis sample. We combine these teachers into a single category for purposes of estimating teacher fixed-effect models.

Table 12 summarizes the number and percentage of observations with students assigned to teachers in each of these categories.

Table 12: Number and percentage of observations with students assigned to "BLOCK," "MISSING," or "MULTIPLE."

		Math		English		Science	
Outcome	**Category**	**N**	**%**	**N**	**%**	**N**	**%**
KY ACT	BLOCK	3,678	3.2	4,091	3.6	4,303	3.8
	MISSING	6,460	5.6	7,663	6.7	26,681	23.3
	MULTIPLE	10,350	9.0	5,722	5.0	11,176	9.8
	Total	**114,465**		**114,465**		**114,465**	
KY PLAN	BLOCK	0	0.0	0	0.0	0	0.0
	MISSING	23,151	28.8	24,705	30.7	29,329	36.4
	MULTIPLE	6,971	8.7	3,579	4.4	5,391	6.7
	Total	**80,490**		**80,490**		**80,490**	

Outcome	Category	Math		English		Science	
		N	%	N	%	N	%
CPS PLAN	No Class	92	0.1	48	0.1	1,023	1.5
	MISSING	1,052	1.5	1,095	1.6	991	1.4
	MULTIPLE	8,332	12.0	10.182	14.6	4,776	6.9
	Small	385	0.6	542	0.8	433	0.6
	Total	**69,741**		**69,741**		**69,741**	

NOTE: Small reflects teachers with at most five students in the sample.

Appendix E. Complete results from all model specifications and outcomes

Table 13: Results for signaling model, mathematics

Kentucky PLAN		(1)	(2)	(3)	(4)	(5)
Effect of having a National Board certified teaching in any semester on student PLAN scores	effect size	0.122	0.096	0.065	0.056	0.070
	std. error	0.034	0.031	0.028	0.024	0.018
	p-value	0.000	0.002	0.018	0.022	0.000
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience proxy		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming EXPLORE		No	No	Yes	Yes	Yes
Observations		80,253	80,253	80,253	80,253	80,253
Schools						338
R^2		0.51	0.51	0.53	0.53	0.55

Kentucky ACT		(1)	(2)	(3)	(4)	(5)
Effect of having a National Board certified teaching in any semester on student ACT scores	effect size	0.099	0.082	0.061	0.056	0.078
	std. error	0.038	0.038	0.024	0.024	0.009
	p-value	0.008	0.030	0.011	0.019	0.000
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience proxy		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming PLAN		No	No	Yes	Yes	Yes
Observations		114,004	114,004	114,004	114,004	114,004
Schools						313
R^2		0.66	0.66	0.68	0.68	0.69

CPS PLAN		(1)	(2)	(3)	(4)	(5)
Effect of having a National Board certified teaching in 9th grade on student PLAN scores	effect size	0.143	0.029	0.072	0.029	0.004
	std. error	0.049	0.028	0.038	0.029	0.027
	p-value	0.003	0.291	0.060	0.318	0.876
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming test score		No	No	Yes	Yes	Yes
Observations		69,741	69,741	69,741	69,741	69,741
Schools						96
R^2		0.58	0.62	0.62	0.63	0.63

CPS ACT		(1)	(2)	(3)	(4)	(5)
Effect of having a National Board certified teaching in 10th or 11th grade on student ACT scores	effect size	0.205	0.103	0.132	0.098	0.077
	std. error	0.027	0.025	0.025	0.024	0.023
	p-value	0.000	0.000	0.000	0.000	0.001
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming test score		No	No	Yes	Yes	Yes
Observations		48,546	48,546	48,546	48,546	48,546
Schools						95
R^2		0.67	0.72	0.71	0.73	0.73

NOTES: Student characteristics include age, number of absences (KY only), racial/ethnic background (black or Hispanic), gender, free and reduced price lunch eligibility, special education and English as a Second Language (ESL) status (KY only), and missing variable indicators. School characteristics include school size (in logs), student-teacher ratio, racial/ethnic composition of student body (percentage of students who are black, percentage of students who are Hispanic), percentage of students eligible for free- and reduced-price lunch, student-administrator ratio and per-pupil spending at the district-level, urban-centric locale indicator (urban, suburban, rural, or town), and school-level average PLAN math, English, and science scores. For Kentucky, the teacher experience proxy is the number of years the teacher appears in the dataset. Standard errors are clustered by teacher.

Table 14: Results for signaling model, English

Kentucky PLAN		(1)	(2)	(3)	(4)	(5)
Effect of having a National Board certified teaching in any semester on student PLAN scores	effect size	-0.004	0.001	-0.010	-0.002	0.000
	std. error	0.025	0.026	0.020	0.020	0.017
	p-value	0.859	0.959	0.606	0.937	0.996
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience proxy		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming EXPLORE		No	No	Yes	Yes	Yes
Observations		80,263	80,263	80,263	80,263	80,263
Schools						338
R^2		0.61	0.61	0.62	0.62	0.63

Kentucky ACT		(1)	(2)	(3)	(4)	(5)
Effect of having a National Board certified teaching in any semester on student ACT scores	effect size	0.076	0.064	0.053	0.028	0.026
	std. error	0.032	0.029	0.022	0.019	0.016
	p-value	0.017	0.027	0.019	0.153	0.098
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience proxy		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming PLAN		No	No	Yes	Yes	Yes
Observations		114,019	114,019	114,019	114,019	114,019
Schools						313
R^2		0.70	0.70	0.71	0.71	0.71

CPS PLAN		(1)	(2)	(3)	(4)	(5)
Effect of having a National Board certified teaching in any semester on student PLAN scores	effect size	0.079	0.043	0.055	0.039	0.056
	std. error	0.031	0.028	0.029	0.027	0.025
	p-value	0.012	0.128	0.055	0.157	0.026
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming test score		No	No	Yes	Yes	Yes
observations		69,741	69,741	69,741	69,741	69,741
schools						96
R^2		0.68	0.70	0.69	0.70	0.70

CPS ACT		(1)	(2)	(3)	(4)	(5)
Effect of having a National Board certified teaching in any semester on student ACT scores	effect size	0.116	0.052	0.063	0.046	0.062
	std. error	0.021	0.015	0.021	0.017	0.017
	p-value	0.000	0.000	0.003	0.006	0.000
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming test score		No	No	Yes	Yes	Yes
Observations		48,546	48,546	48,546	48,546	48,546
Schools						95
R^2		0.71	0.74	0.73	0.74	0.74

NOTES: Student characteristics include age, number of absences (KY only), racial/ethnic background (black or Hispanic), gender, free and reduced price lunch eligibility, special education and English as a Second Language (ESL) status (KY only), and missing variable indicators. School characteristics include school size (in logs), student-teacher ratio, racial/ethnic composition of student body (percentage of students who are black, percentage of students who are Hispanic), percentage of students eligible for free- and reduced-price lunch, student-administrator ratio and per-pupil spending at the district-level, urban-centric locale indicator (urban, suburban, rural, or town), and school-level average PLAN math, English, and science scores. For Kentucky, the teacher experience proxy is the number of years the teacher appears in the dataset. Standard errors are clustered by teacher.

Table 15: Results for signaling model, science

Kentucky PLAN		(1)	(2)	(3)	(4)	(5)
Effect of having a National Board certified teaching in any semester on student PLAN scores	effect size	0.032	0.022	0.008	0.005	-0.015
	std. error	0.028	0.025	0.032	0.027	0.026
	p-value	0.245	0.365	0.807	0.843	0.555
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience proxy		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming EXPLORE		No	No	Yes	Yes	Yes
Observations		80,163	80,163	80,163	80,163	80,163
Schools						338
R^2		0.42	0.43	0.43	0.43	0.44

Kentucky ACT		(1)	(2)	(3)	(4)	(5)
Effect of having a National Board certified teaching in any semester on student ACT scores	effect size	0.040	0.022	0.021	0.006	0.026
	std. error	0.038	0.042	0.034	0.038	0.030
	p-value	0.291	0.591	0.538	0.866	0.388
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience proxy		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming PLAN		No	No	Yes	Yes	Yes
Observations		113,923	113,923	113,923	113,923	113,923
Schools						313
R^2		0.49	0.50	0.50	0.51	0.52

CPS PLAN		(1)	(2)	(3)	(4)	(5)
Effect of having a National Board certified teaching in any semester on student PLAN scores	effect size	0.304	0.028	0.124	0.023	-0.027
	std. error	0.074	0.031	0.045	0.029	0.035
	p-value	0.000	0.361	0.006	0.427	0.449
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming test score		No	No	Yes	Yes	Yes
Observations		69,741	69,741	69,741	69,741	69,741
Schools						96
R^2		0.49	0.54	0.53	0.55	0.55

CPS ACT		(1)	(2)	(3)	(4)	(5)
Effect of having a National Board certified teaching in any semester on student ACT scores	effect size	0.190	0.020	0.072	0.013	0.013
	std. error	0.033	0.026	0.033	0.023	0.021
	p-value	0.000	0.432	0.027	0.576	0.545
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming test score		No	No	Yes	Yes	Yes
Observations		48,546	48,546	48,546	48,546	48,546
Schools						95
R^2		0.52	0.57	0.56	0.58	0.58

NOTES: Student characteristics include age, number of absences (KY only), racial/ethnic background (black or Hispanic), gender, free and reduced price lunch eligibility, special education and English as a Second Language (ESL) status (KY only), and missing variable indicators. School characteristics include school size (in logs), student-teacher ratio, racial/ethnic composition of student body (percentage of students who are black, percentage of students who are Hispanic), percentage of students eligible for free- and reduced-price lunch, student-administrator ratio and per-pupil spending at the district-level, urban-centric locale indicator (urban, suburban, rural, or town), and school-level average PLAN math, English, and science scores. For Kentucky, the teacher experience proxy is the number of years the teacher appears in the dataset. Standard errors are clustered by teacher.

Table 16: Results for signaling model, all subjects (pooled)

Kentucky PLAN		(1)	(2)	(3)	(4)	(5)
Effect of having a National Board certified teaching in any semester on student PLAN scores	effect size	0.042	0.031	0.015	0.013	0.010
	std. error	0.018	0.017	0.014	0.013	0.011
	p-value	0.017	0.070	0.274	0.336	0.366
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience proxy		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming EXPLORE		No	No	Yes	Yes	Yes
Observations		240,679	240,679	240,679	240,679	240,679
Schools						338
R^2		0.51	0.51	0.52	0.52	0.53

Kentucky ACT		(1)	(2)	(3)	(4)	(5)
Effect of having a National Board certified teaching in any semester on student ACT scores	effect size	0.071	0.058	0.042	0.034	0.038
	std. error	0.022	0.022	0.015	0.015	0.012
	p-value	0.001	0.008	0.005	0.025	0.002
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience proxy		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming PLAN		No	No	Yes	Yes	Yes
Observations		341,946	341,946	341,946	341,946	341,946
Schools						313
R^2		0.61	0.61	0.62	0.62	0.62

CPS PLAN		(1)	(2)	(3)	(4)	(5)
Effect of having a National Board certified teaching in any semester on student PLAN scores	effect size	0.167	0.035	0.080	0.030	0.019
	std. error	0.036	0.021	0.029	0.021	0.021
	p-value	0.000	0.088	0.005	0.146	0.378
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming test score		No	No	Yes	Yes	Yes
observations		209,223	209,223	209,223	209,223	209,223
schools						96
R^2		0.58	0.62	0.61	0.62	0.62

CPS ACT		(1)	(2)	(3)	(4)	(5)
Effect of having a National Board certified teaching in any semester on student ACT scores	effect size	0.163	0.062	0.087	0.056	0.054
	std. error	0.017	0.012	0.022	0.012	0.012
	p-value	0.000	0.000	0.000	0.000	0.000
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming test score		No	No	Yes	Yes	Yes
Observations		145,638	145,638	145,638	145,638	145,638
Schools						95
R^2		0.63	0.67	0.66	0.67	0.67

NOTES: Student characteristics include age, number of absences (KY only), racial/ethnic background (black or Hispanic), gender, free and reduced price lunch eligibility, special education and English as a Second Language (ESL) status (KY only), and missing variable indicators. School characteristics include school size (in logs), student-teacher ratio, racial/ethnic composition of student body (percentage of students who are black, percentage of students who are Hispanic), percentage of students eligible for free- and reduced-price lunch, student-administrator ratio and per-pupil spending at the district-level, urban-centric locale indicator (urban, suburban, rural, or town), and school-level average PLAN math, English, and science scores. For Kentucky, the teacher experience proxy is the number of years the teacher appears in the dataset. Standard errors are clustered by teacher.

Table 17: Results for screening model, mathematics

Kentucky PLAN		(1)	(2)	(3)	(4)	(5)
Effect of variable for number of semesters with an ever-certified teacher on PLAN scores	effect size	0.071	0.061	0.043	0.037	0.043
	std. error	0.016	0.015	0.013	0.012	0.010
	p-value	0.000	0.000	0.001	0.003	0.000
Effect of variable for number of semesters with a never-certified teacher on PLAN scores	effect size	-0.014	-0.020	0.001	-0.002	-0.010
	std. error	0.026	0.028	0.023	0.024	0.021
	p-value	0.607	0.470	0.954	0.944	0.637
Effect of variable for number of semesters with an ever-withdrawn teacher on PLAN scores	effect size	-0.022	-0.024	-0.023	-0.021	0.014
	std. error	0.018	0.018	0.018	0.017	0.020
	p-value	0.212	0.180	0.199	0.230	0.488
Test: Ever certified - never certified	effect size	0.085	0.081	0.041	0.039	0.053
	std. error	0.029	0.032	0.022	0.024	0.022
	p-value	0.004	0.012	0.060	0.115	0.014
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience proxy		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming EXPLORE		No	No	Yes	Yes	Yes
Observations		80,253	80,253	80,253	80,253	80,253
Schools						338
R^2		0.51	0.51	0.53	0.53	0.55

Kentucky ACT		(1)	(2)	(3)	(4)	(5)
Effect of variable for number of semesters with an ever-certified teacher on ACT scores	effect size	0.053	0.045	0.028	0.026	0.039
	std. error	0.014	0.014	0.009	0.009	0.008
	p-value	0.000	0.001	0.001	0.003	0.000
Effect of variable for number of semesters with a never-certified teacher on ACT scores	effect size	0.031	0.015	0.021	0.020	0.003
	std. error	0.019	0.014	0.009	0.010	0.012
	p-value	0.105	0.277	0.023	0.050	0.822
Effect of variable for number of semesters with an ever-withdrawn teacher on ACT scores	effect size	0.040	0.057	0.041	0.045	0.057
	std. error	0.029	0.030	0.026	0.025	0.020
	p-value	0.169	0.056	0.107	0.070	0.005
Test: Ever certified - never certified	effect size	0.022	0.030	0.006	0.005	0.036
	std. error	0.026	0.022	0.014	0.014	0.014
	p-value	0.395	0.175	0.652	0.707	0.011
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience proxy		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming PLAN		No	No	Yes	Yes	Yes
Observations		114,004	114,004	114,004	114,004	114,004
Schools						313
R^2		0.66	0.66	0.68	0.68	0.69

CPS PLAN		(1)	(2)	(3)	(4)	(5)
Effect of variable for number of years with an ever-certified teacher on PLAN scores	effect size	0.120	0.041	0.062	0.036	0.027
	std. error	0.043	0.024	0.035	0.026	0.026
	p-value	0.005	0.091	0.071	0.163	0.299
Effect of variable for number of years with a never-certified teacher on PLAN scores	effect size	-0.018	-0.005	0.022	0.013	0.007
	std. error	0.036	0.029	0.033	0.029	0.032
	p-value	0.620	0.856	0.516	0.659	0.835
Effect of variable for number of years with an outcome unknown teacher on PLAN scores	effect size	0.085	0.043	0.058	0.040	0.064
	std. error	0.067	0.055	0.060	0.057	0.050
	p-value	0.205	0.431	0.335	0.480	0.202
Test: Ever certified - never certified	effect size	0.138	0.046	0.041	0.023	0.021
	std. error	0.052	0.036	0.039	0.035	0.037
	p-value	0.008	0.204	0.297	0.502	0.579
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming test score		No	No	Yes	Yes	Yes
Observations		69,741	69,741	69,741	69,741	69,741
Schools						96
R^2		0.58	0.62	0.62	0.63	0.63

CPS ACT		(1)	(2)	(3)	(4)	(5)
Effect of variable for number of years with an ever-certified teacher on ACT scores	effect size	0.169	0.090	0.161	0.081	0.086
	std. error	0.020	0.018	0.021	0.022	0.018
	p-value	0.000	0.000	0.000	0.000	0.000
Effect of variable for number of years with a never-certified teacher on ACT scores	effect size	0.007	0.008	0.005	0.008	0.014
	std. error	0.025	0.026	0.024	0.026	0.023
	p-value	0.786	0.764	0.838	0.760	0.545
Effect of variable for number of years with an outcome unknown teacher on ACT scores	effect size	0.070	0.057	0.067	0.057	0.059
	std. error	0.025	0.015	0.023	0.015	0.014
	p-value	0.005	0.000	0.004	0.000	0.000
Test: Ever certified - never certified	effect size	0.162	0.082	0.157	0.073	0.072
	std. error	0.031	0.031	0.031	0.034	0.030
	p-value	0.000	0.009	0.000	0.031	0.016
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming PLAN		No	No	Yes	Yes	Yes
Observations		48,546	48,546	48,546	48,546	48,546
Schools						95

NOTES: Student characteristics include age, number of absences (KY only), racial/ethnic background (black or Hispanic), gender, free and reduced price lunch eligibility, special education and English as a Second Language (ESL) status (KY only), and missing variable indicators. School characteristics include school size (in logs), student-teacher ratio, racial/ethnic composition of student body (percentage of students who are black, percentage of students who are Hispanic), percentage of students eligible for free- and reduced-price lunch, student-administrator ratio and per-pupil spending at the district-level, urban-centric locale indicator (urban, suburban, rural, or town), and school-level average PLAN math, English, and science scores. For Kentucky, the teacher experience proxy is the number of years the teacher appears in the dataset. Standard errors are clustered by teacher.

Table 18: Results for screening model, English

Kentucky PLAN		(1)	(2)	(3)	(4)	(5)
Effect of variable for number of semesters with an ever-certified teacher on PLAN scores	effect size	-0.007	-0.002	-0.007	-0.002	0.002
	std. error	0.012	0.012	0.009	0.009	0.008
	p-value	0.590	0.853	0.441	0.808	0.836
Effect of variable for number of semesters with a never-certified teacher on PLAN scores	effect size	0.018	0.017	0.015	0.017	0.016
	std. error	0.011	0.011	0.011	0.011	0.009
	p-value	0.087	0.136	0.166	0.128	0.069
Effect of variable for number of semesters with an ever-withdrawn teacher on PLAN scores	effect size	-0.046	-0.036	-0.033	-0.029	-0.019
	std. error	0.022	0.016	0.020	0.018	0.017
	p-value	0.039	0.013	0.089	0.108	0.251
Test: Ever certified - never certified	effect size	-0.025	-0.019	-0.023	-0.019	-0.014
	std. error	0.016	0.016	0.014	0.014	0.012
	p-value	0.133	0.242	0.112	0.171	0.223
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience proxy		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming EXPLORE		No	No	Yes	Yes	Yes
observations		80,263	80,263	80,263	80,263	80,263
schools						338
R^2		0.61	0.61	0.62	0.62	0.63

Kentucky ACT		(1)	(2)	(3)	(4)	(5)
Effect of variable for number of semesters with an ever-certified teacher on ACT scores	effect size	0.024	0.015	0.014	0.003	0.012
	std. error	0.012	0.011	0.010	0.008	0.007
	p-value	0.046	0.190	0.156	0.682	0.105
Effect of variable for number of semesters with a never-certified teacher on ACT scores	effect size	0.002	-0.007	0.003	-0.006	-0.001
	std. error	0.018	0.019	0.014	0.014	0.012
	p-value	0.905	0.717	0.824	0.667	0.917
Effect of variable for number of semesters with an ever-withdrawn teacher on ACT scores	effect size	-0.035	-0.026	-0.034	-0.026	0.004
	std. error	0.014	0.013	0.011	0.011	0.017
	p-value	0.012	0.043	0.002	0.024	0.805
Test: Ever certified - never certified	effect size	0.022	0.022	0.011	0.009	0.013
	std. error	0.021	0.022	0.017	0.017	0.015
	p-value	0.296	0.315	0.519	0.568	0.391
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience proxy		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming PLAN		No	No	Yes	Yes	Yes
observations		114,019	114,019	114,019	114,019	114,019
schools						313
R^2		0.70	0.70	0.71	0.71	0.71

CPS PLAN		(1)	(2)	(3)	(4)	(5)
Effect of variable for number of years with an ever-certified teacher on PLAN scores	effect size	0.094	0.036	0.052	0.030	0.045
	std. error	0.028	0.019	0.026	0.020	0.017
	p-value	0.001	0.065	0.046	0.129	0.009
Effect of variable for number of years with a never-certified teacher on PLAN scores	effect size	0.027	0.016	0.056	0.029	-0.010
	std. error	0.030	0.027	0.029	0.028	0.024
	p-value	0.372	0.550	0.052	0.288	0.667
Effect of variable for number of years with an outcome unknown teacher on PLAN scores	effect size	0.049	0.014	0.042	0.017	0.024
	std. error	0.032	0.021	0.027	0.020	0.018
	p-value	0.129	0.516	0.112	0.399	0.188
Test: Ever certified - never certified	effect size	0.067	0.019	-0.004	0.001	0.056
	std. error	0.036	0.031	0.029	0.030	0.027
	p-value	0.062	0.525	0.899	0.974	0.043
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience proxy		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming test score		No	No	Yes	Yes	Yes
Observations		69,741	69,741	69,741	69,741	69,741
Schools						96
R^2		0.68	0.70	0.69	0.70	0.70

CPS ACT		(1)	(2)	(3)	(4)	(5)
Effect of variable for number of years with an ever-certified teacher on ACT scores	effect size	0.082	0.034	0.039	0.027	0.046
	std. error	0.016	0.011	0.016	0.012	0.015
	p-value	0.000	0.002	0.013	0.028	0.002
Effect of variable for number of years with a never-certified teacher on ACT scores	effect size	0.029	0.002	0.038	0.009	0.005
	std. error	0.036	0.027	0.026	0.025	0.018
	p-value	0.428	0.936	0.148	0.725	0.791
Effect of variable for number of years with an outcome unknown teacher on ACT scores	effect size	0.008	0.012	0.019	0.014	0.005
	std. error	0.019	0.014	0.017	0.014	0.013
	p-value	0.662	0.370	0.263	0.317	0.687
Test: Ever certified - never certified	effect size	0.053	0.032	0.001	0.018	0.041
	std. error	0.038	0.027	0.026	0.026	0.021
	p-value	0.157	0.249	0.957	0.477	0.049
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming PLAN		No	No	Yes	Yes	Yes
Observations		48,546	48,546	48,546	48,546	48,546
Schools						95
R^2		0.71	0.74	0.73	0.74	0.74

NOTES: Student characteristics include age, number of absences (KY only), racial/ethnic background (black or Hispanic), gender, free and reduced price lunch eligibility, special education and English as a Second Language (ESL) status (KY only), and missing variable indicators. School characteristics include school size (in logs), student-teacher ratio, racial/ethnic composition of student body (percentage of students who are black, percentage of students who are Hispanic), percentage of students eligible for free- and reduced-price lunch, student-administrator ratio and per-pupil spending at the district-level, urban-centric locale indicator (urban, suburban, rural, or town), and school-level average PLAN math, English, and science scores. For Kentucky, the teacher experience proxy is the number of years the teacher appears in the dataset. Standard errors are clustered by teacher.

Table 19: Results for screening model, science

Kentucky PLAN		(1)	(2)	(3)	(4)	(5)
Effect of variable for number of semesters with an ever-certified teacher on PLAN scores	effect size	0.033	0.027	0.023	0.021	0.000
	std. error	0.015	0.014	0.018	0.015	0.013
	p-value	0.029	0.052	0.190	0.157	0.999
Effect of variable for number of semesters with a never-certified teacher on PLAN scores	effect size	-0.003	0.002	0.008	0.010	-0.006
	std. error	0.014	0.012	0.014	0.012	0.013
	p-value	0.803	0.883	0.583	0.398	0.643
Effect of variable for number of semesters with an ever-withdrawn teacher on PLAN scores	effect size	0.003	-0.005	0.013	0.008	-0.002
	std. error	0.027	0.025	0.021	0.020	0.018
	p-value	0.925	0.858	0.530	0.680	0.924
Test: Ever certified - never certified	effect size	0.037	0.025	0.016	0.011	0.006
	std. error	0.024	0.021	0.028	0.023	0.020
	p-value	0.119	0.226	0.573	0.634	0.753
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience proxy		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming EXPLORE		No	No	Yes	Yes	Yes
observations		80,163	80,163	80,163	80,163	80,163
schools						338
R^2		0.42	0.43	0.43	0.43	0.44

Kentucky ACT		(1)	(2)	(3)	(4)	(5)
Effect of variable for number of semesters with an ever-certified teacher on ACT scores	effect size	0.011	0.005	0.005	-0.001	0.011
	std. error	0.014	0.015	0.014	0.014	0.012
	p-value	0.431	0.730	0.696	0.971	0.356
Effect of variable for number of semesters with a never-certified teacher on ACT scores	effect size	-0.002	0.000	-0.005	-0.003	-0.008
	std. error	0.022	0.018	0.019	0.015	0.012
	p-value	0.926	0.999	0.797	0.817	0.499
Effect of variable for number of semesters with an ever-withdrawn teacher on ACT scores	effect size	-0.010	-0.066	-0.030	-0.074	-0.039
	std. error	0.024	0.022	0.022	0.020	0.019
	p-value	0.696	0.003	0.178	0.000	0.044
Test: Ever certified - never certified	effect size	0.013	0.005	0.010	0.003	0.019
	std. error	0.026	0.023	0.023	0.020	0.017
	p-value	0.606	0.820	0.655	0.886	0.247
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience proxy		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming PLAN		No	No	Yes	Yes	Yes
observations		113,923	113,923	113,923	113,923	113,923
schools						313
R^2		0.49	0.50	0.50	0.51	0.52

CPS PLAN		(1)	(2)	(3)	(4)	(5)
Effect of variable for number of years with an ever-certified teacher on PLAN scores	effect size	0.281	0.040	0.109	0.030	-0.010
	std. error	0.063	0.026	0.041	0.025	0.029
	p-value	0.000	0.124	0.008	0.243	0.725
Effect of variable for number of years with a never-certified teacher on PLAN scores	effect size	0.041	0.011	0.055	0.026	0.020
	std. error	0.044	0.028	0.028	0.027	0.034
	p-value	0.347	0.691	0.047	0.341	0.547
Effect of variable for number of years with an outcome unknown teacher on PLAN scores	effect size	0.087	0.031	0.034	0.021	0.040
	std. error	0.051	0.022	0.038	0.025	0.022
	p-value	0.088	0.164	0.376	0.364	0.070
Test: Ever certified - never certified	effect size	0.240	0.029	0.054	0.004	-0.030
	std. error	0.073	0.034	0.040	0.032	0.040
	p-value	0.001	0.396	0.180	0.911	0.447
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming test score		No	No	Yes	Yes	Yes
Observations		69,741	69,741	69,741	69,741	69,741
Schools						96
R^2		0.50	0.54	0.53	0.55	0.55

CPS ACT		(1)	(2)	(3)	(4)	(5)
Effect of variable for number of years with an ever-certified teacher on ACT scores	effect size	0.148	0.015	0.054	0.009	0.016
	std. error	0.024	0.016	0.026	0.015	0.016
	p-value	0.000	0.347	0.035	0.528	0.315
Effect of variable for number of years with a never-certified teacher on ACT scores	effect size	0.056	-0.008	0.026	-0.009	-0.019
	std. error	0.026	0.015	0.013	0.014	0.016
	p-value	0.034	0.576	0.053	0.548	0.250
Effect of variable for number of years with an outcome unknown teacher on ACT scores	effect size	0.049	0.046	0.031	0.039	0.050
	std. error	0.026	0.019	0.015	0.019	0.019
	p-value	0.058	0.014	0.041	0.040	0.010
Test: Ever certified - never certified	effect size	0.092	0.024	0.029	0.018	0.035
	std. error	0.035	0.021	0.022	0.020	0.023
	p-value	0.009	0.258	0.200	0.365	0.128
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming PLAN		No	No	Yes	Yes	Yes
Observations		48,546	48,546	48,546	48,546	48,546
Schools						95
R^2		0.52	0.57	0.56	0.58	0.58

NOTES: Student characteristics include age, number of absences (KY only), racial/ethnic background (black or Hispanic), gender, free and reduced price lunch eligibility, special education and English as a Second Language (ESL) status (KY only), and missing variable indicators. School characteristics include school size (in logs), student-teacher ratio, racial/ethnic composition of student body (percentage of students who are black, percentage of students who are Hispanic), percentage of students eligible for free- and reduced-price lunch, student-administrator ratio and per-pupil spending at the district-level, urban-centric locale indicator (urban, suburban, rural, or town), and school-level average PLAN math, English, and science scores. For Kentucky, the teacher experience proxy is the number of years the teacher appears in the dataset. Standard errors are clustered by teacher.

Table 20: Results for screening model, all subjects

Kentucky PLAN		(1)	(2)	(3)	(4)	(5)
Effect of variable for number of semesters with an ever-certified teacher on PLAN scores	effect size	0.027	0.022	0.014	0.013	0.010
	std. error	0.009	0.009	0.008	0.008	0.007
	p-value	0.003	0.015	0.082	0.091	0.152
Effect of variable for number of semesters with a never-certified teacher on PLAN scores	effect size	0.003	0.003	0.009	0.009	0.003
	std. error	0.008	0.009	0.008	0.008	0.007
	p-value	0.694	0.721	0.233	0.244	0.643
Effect of variable for number of semesters with an ever-withdrawn teacher on PLAN scores	effect size	-0.022	-0.022	-0.013	-0.013	-0.017
	std. error	0.013	0.013	0.012	0.012	0.012
	p-value	0.096	0.089	0.256	0.253	0.158
Test: Ever certified - never certified	effect size	0.024	0.019	0.005	0.004	0.007
	std. error	0.013	0.013	0.013	0.012	0.011
	p-value	0.063	0.139	0.692	0.767	0.541
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience proxy		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming EXPLORE		No	No	Yes	Yes	Yes
observations		240,679	240,679	240,679	240,679	240,679
schools						338
R^2		0.51	0.51	0.52	0.52	0.53

Kentucky ACT		(1)	(2)	(3)	(4)	(5)
Effect of variable for number of semesters with an ever-certified teacher on ACT scores	effect size	0.031	0.022	0.015	0.010	0.015
	std. error	0.008	0.008	0.006	0.006	0.005
	p-value	0.000	0.006	0.013	0.091	0.006
Effect of variable for number of semesters with a never-certified teacher on ACT scores	effect size	0.007	-0.002	0.005	-0.002	-0.005
	std. error	0.013	0.011	0.010	0.010	0.009
	p-value	0.609	0.876	0.645	0.832	0.588
Effect of variable for number of semesters with an ever-withdrawn teacher on ACT scores	effect size	-0.008	-0.008	-0.011	-0.009	0.005
	std. error	0.014	0.013	0.012	0.011	0.010
	p-value	0.557	0.549	0.350	0.415	0.627
Test: Ever certified - never certified	effect size	0.024	0.024	0.010	0.012	0.020
	std. error	0.015	0.014	0.012	0.012	0.010
	p-value	0.117	0.082	0.401	0.303	0.048
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience proxy		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming PLAN		No	No	Yes	Yes	Yes
observations		341,946	341,946	341,946	341,946	341,946
schools						313
R^2		0.61	0.61	0.62	0.62	0.62

CPS PLAN		(1)	(2)	(3)	(4)	(5)
Effect of variable for number of years with an ever-certified teacher on PLAN scores	effect size	0.161	0.044	0.076	0.036	0.029
	std. error	0.032	0.017	0.029	0.019	0.018
	p-value	0.000	0.010	0.008	0.050	0.112
Effect of variable for number of years with a never-certified teacher on PLAN scores	effect size	0.017	0.005	0.045	0.020	0.014
	std. error	0.024	0.018	0.023	0.018	0.019
	p-value	0.487	0.776	0.055	0.281	0.475
Effect of variable for number of years with an outcome unknown teacher on PLAN scores	effect size	0.067	0.024	0.042	0.022	0.035
	std. error	0.031	0.020	0.029	0.021	0.020
	p-value	0.031	0.232	0.145	0.286	0.076
Test: Ever certified - never certified	effect size	0.144	0.038	0.031	0.016	0.015
	std. error	0.034	0.022	0.023	0.021	0.021
	p-value	0.000	0.082	0.174	0.444	0.468
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming test score		No	No	Yes	Yes	Yes
Observations		209,223	209,223	209,223	209,223	209,223
Schools						96
R^2		0.58	0.62	0.61	0.62	0.62

CPS ACT		(1)	(2)	(3)	(4)	(5)
Effect of variable for number of years with an ever-certified teacher on ACT scores	effect size	0.126	0.050	0.066	0.044	0.046
	std. error	0.013	0.009	0.016	0.009	0.009
	p-value	0.000	0.000	0.000	0.000	0.000
Effect of variable for number of years with a never-certified teacher on ACT scores	effect size	0.034	-0.003	0.024	-0.001	0.002
	std. error	0.018	0.012	0.014	0.011	0.011
	p-value	0.068	0.783	0.080	0.960	0.868
Effect of variable for number of years with an outcome unknown teacher on ACT scores	effect size	0.039	0.035	0.036	0.034	0.037
	std. error	0.013	0.009	0.010	0.009	0.009
	p-value	0.004	0.000	0.000	0.000	0.000
Test: Ever certified - never certified	effect size	0.093	0.054	0.042	0.045	0.045
	std. error	0.021	0.014	0.014	0.014	0.014
	p-value	0.000	0.000	0.002	0.001	0.001
Additional controls:						
Student characteristics		Yes	Yes	Yes	Yes	Yes
Teacher experience		Yes	Yes	Yes	Yes	Yes
School characteristics		No	Yes	No	Yes	No
School FE		No	No	No	No	Yes
Average incoming PLAN		No	No	Yes	Yes	Yes
Observations		145,638	145,638	145,638	145,638	145,638
Schools						95
R^2		0.63	0.67	0.66	0.67	0.67

NOTES: Student characteristics include age, number of absences (KY only), racial/ethnic background (black or Hispanic), gender, free and reduced price lunch eligibility, special education and English as a Second Language (ESL) status (KY only), and missing variable indicators. School characteristics include school size (in logs), student-teacher ratio, racial/ethnic composition of student body (percentage of students who are black, percentage of students who are Hispanic), percentage of students eligible for free- and reduced-price lunch, student-administrator ratio and per-pupil spending at the district-level, urban-centric locale indicator (urban, suburban, rural, or town), and school-level average PLAN math, English, and science scores. For Kentucky, the teacher experience proxy is the number of years the teacher appears in the dataset. Standard errors are clustered by teacher.

Table 21: Results for human capital model, all subjects (pooled)

Kentucky ACT		(1)	(2)	(4)	(5)
Current applicants	effect size	0.061	0.043	0.019	0.042
	std. error	0.074	0.060	0.056	0.080
	p-value	0.408	0.473	0.737	0.595
Past applicants	effect size	-0.003	-0.004	-0.018	-0.018
	std. error	0.038	0.042	0.046	0.055
	p-value	0.934	0.922	0.698	0.739
Additional controls:					
Student characteristics		Yes	Yes	Yes	Yes
Teacher experience proxy		Yes	Yes	Yes	Yes
School characteristics		No	Yes	Yes	No
Teacher FE		Yes	Yes	Yes	Yes
School FE		No	No	No	Yes
Average incoming PLAN		No	No	Yes	Yes
observations		342,462	342,462	342,462	342,462
schools					313
teachers		5,438	5,438	5,438	5,438
R^2		0.59	0.60	0.60	0.59

NOTES: Student covariates include prior test score, demographic variables; model includes teacher fixed effects for current teacher and school-level fixed effects. The omitted group is future applicants – teachers who have not applied but will in the future. Standard errors are clustered by teacher.

CPS PLAN		(1)	(2)	(3)	(4)
Current applicants	effect size	0.023	0.029	0.019	0.023
	std. error	0.023	0.023	0.023	0.022
	p-value	0.321	0.196	0.427	0.307
Past applicants	effect size	0.017	-0.002	-0.002	-0.008
	std. error	0.025	0.025	0.025	0.025
	p-value	0.511	0.922	0.951	0.751
Additional controls:					
Student characteristics		Yes	Yes	Yes	Yes
Teacher experience		Yes	Yes	Yes	Yes
School characteristics		No	Yes	Yes	No
Teacher FE		Yes	Yes	Yes	Yes
School FE		No	No	No	Yes
Average incoming test score		No	No	Yes	Yes
observations		209,223	209,223	209,223	209,223
schools					99
teachers		2,360	2,360	2,360	2,360
R^2		0.64	0.64	0.64	0.64

CPS ACT		(1)	(2)	(3)	(4)
Grade 10 teacher					
Current applicants	effect size	0.019	-0.006	0.021	0.004
	std. error	0.021	0.021	0.021	0.022
	p-value	0.371	0.763	0.314	0.855
Past applicants	effect size	-0.012	-0.038	-0.016	-0.026
	std. error	0.027	0.027	0.027	0.026
	p-value	0.648	0.159	0.554	0.312
Grade 11 teacher					
Current applicants	effect size	0.005	-0.005	0.000	0.002
	std. error	0.023	0.024	0.025	0.024
	p-value	0.832	0.839	0.986	0.929
Past applicants	effect size	-0.004	-0.025	-0.017	-0.020
	std. error	0.026	0.025	0.027	0.026
	p-value	0.887	0.324	0.514	0.436
Additional controls:					
Student characteristics		Yes	Yes	Yes	Yes
Teacher experience		Yes	Yes	Yes	Yes
School characteristics		No	Yes	Yes	No
Teacher FE		Yes	Yes	Yes	Yes
School FE		No	No	No	Yes
Average incoming test score		No	No	Yes	Yes
Observations		143,898	143,898	143,898	143898
Schools					94
Teachers		2,856	2,856	2,856	2,856
R^2		0.70			0.70

Appendix F. Analysis of ceiling effects on instructional improvement

In order to determine whether teachers with lower scores for instructional quality at baseline grew more between baseline and subsequent observations, we created a single vector measuring standardized score change on each subscale. Next, we ran a regression model that included a dichotomous variable for whether the teacher was a National Board applicant (1=yes, 0=no), and a series of variables to indicate the quartile of the teacher's rating at baseline. Interaction terms were added for the National Board–applicant variable and the quartile variables to test whether applicants in the bottom quartile of ratings at baseline experienced more growth in ratings than did applicants in the top quartile of ratings at baseline. We also included control variables for each of the subscales and the time point of the observation. The model included robust standard errors, clustered on teacher.

The regression results indicate no statistically significant effect for the National Board–applicant variable, or for any of the interaction terms between the applicant variable and the quartile of baseline performance.

This page intentionally left blank

Glossary

CPS	Chicago Public Schools district
EPAS	Educational Planning and Assessment System
ESL	English as a Second Language
FRL	free or reduced-price lunch
IEP	Individualized Education Program
KDE	Kentucky Department of Education
LBD	Leadership by Design (classroom observation instrument)
NAEP	National Assessment of Educational Progress
NBC	National Board certification
NBCT	National Board–certified teacher
NBPTS	National Board for Professional Teaching Standards
NCLB	No Child Left Behind
SY	school year

This page intentionally left blank

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics, 25*(1), 95–135.

Burke, D., Cavalluzzo, L., Hansen, M., Harris, D., Lien, J., & Wenger, J. (2004). *Critical issues in recruiting and retaining the mathematics and science workforce: Summary of research in Miami-Dade County Public Schools.* Alexandria, VA: The CNA Corporation.

Cantrell, S., Fullerton, J., Kane, T. J., & Staiger, D. O. (2008). *National board certification and teacher effectiveness: Evidence from a random assignment experiment* (NBER Working Paper No. 14608). Cambridge, MA: National Bureau of Economic Research.

Carnegie Task Force on Teaching as a Profession. (1986). *A nation prepared: Teachers for the 21st century.* New York: Carnegie Corporation of New York, Forum on Education and the Economy.

Cavalluzzo, L. (2004). *Is National Board certification an effective signal of teacher quality?* Alexandria, VA: The CNA Corporation.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources, 41*(4), 778–820.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). *How and why do teacher credentials matter for student achievement?* (NBER Working Paper No. 12828). Cambridge, MA: National Bureau of Economic Research.

Cohen, C., & Rice, J. K. (2005). *National Board certification as professional development: Pathways to success.* Washington, DC: The Finance Project.

Darling-Hammond, L., Atkin, J. M., Sato, M., Chung, R. R., Dean, S., & Greenwald, E. (2007). *Influences of National Board certification on teachers' classroom assessment practices.* Unpublished paper, Stanford University.

Eide, E. R., & Showalter, M. H. (2010). Human capital. In D. J. Brewer & P. J. McEwan (Eds.), *Economics of education* (pp. 27–32). Amsterdam: Elsevier.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915–945.

Goldhaber, D. D. (2002). The mystery of good teaching. *Education Next, 2*(1), 50–55.

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

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National Board certification as a signal of effective teaching. *The Review of Economics and Statistics, 89*(1), 134–150.

Hakel, M. D., Koenig, J. A., & Elliott, S. W. (Eds.). (2008). *Assessing accomplished teaching: Advanced-level certification programs.* Washington, DC: National Academies Press.

Harris, D. N., & Sass, T. R. (2006). *The effects of teacher training on teacher value-added.* Florida State University. Unpublished manuscript.

Harris, D. N., & Sass, T. R. (2007). *Teacher training, teacher quality, and student achievement.* Washington, DC: The Urban Institute, National Center for Analysis of Longitudinal Data in Education Research (CALDER).

Indiana Professional Standards Board. (2002). *Status of National Board certified teachers in Indiana.* Retrieved Feb. 28, 2014, from http://www.nbpts.org/resources/research/browse_studies?ID=26

Jacob, B. A., & Lefgren, L. (2004). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. *The Journal of Human Resources, 50*(1), 50–79.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review, 27,* 615–631.

McCaffrey, D. F., & Rivkin, S. (2007). *Empirical investigations of the effects of National Board of Professional Teaching Standards certified teachers on student outcomes.* Unpublished manuscript.

National Board for Professional Teaching Standards. (2013). *Number of Board-certified teachers tops 100,000 as new studies show significant gains in student achievement.* Retrieved from http://www.nbpts.org/nbpts-2013-01-08-nbct-student-achievement

National Board for Professional Teaching Standards. (2012). *NBCT certification renewal: Profile of professional growth.* Retrieved from http://www.nbpts.org/sites/default/files/documents/Candidate-Center/Profile_ProfGrowth_20130408.pdf

National Board for Professional Teaching Standards. (2011). *2011 guide to National Board certification.* Arlington, VA.

National Board for Professional Teaching Standards. (2002). *What teachers should know and be able to do.* Retrieved from http://www.nbpts.org/sites/default/files/documents/certificates/what_teachers_should_know.pdf

National Research Council. (2008). *Assessing accomplished teaching, advanced level certification programs.* Retrieved from http://www.nap.edu/catalog.php?record_id=12224

Podgursky, M., Springer, M. G., & Hutton, R.(2010). Teacher training and preparation in the United States. In D. J. Brewer & P. J. McEwan (Eds.), *Economics of education* (pp. 290–295). Amsterdam: Elsevier.

Rampey, B. D., Dion, G. S., & Donahue, P. L. (2009). *NAEP 2008. Trends in academic progress* (NCES 2009–479). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools and academic achievement. *Econometrica, 73*(2), 417–458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2), 247–252.

Sanders, W. L., Ashton, J. J., & Wright, S. P. (2005). *Comparison of the effects of NBPTS certified teachers with other teachers on the rate of student academic progress. Final report.* Arlington, VA: National Board for Professional Teaching Standards.

Smylie, M. A., Allensworth, E., Greenberg, R. C., Harris, R., & Luppescu, S. (2001). *Teacher professional development in Chicago: Supporting effective practice.* Chicago, IL: Consortium on Chicago School Research.

Yankelovich Partners. (2001). *Leading from the classroom: Highlights from the 2001 NBPTS National Board Certified Teacher Leadership Survey.* Retrieved from http://www.nbpts.org/UserFiles/File/leading_from_the_classroom(1).pdf

Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement.* Washington, DC: National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

List of figures

List of tables


CNA
ANALYSIS & SOLUTIONS