EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Volume 12 Number 46 **September 8, 2004** **ISSN 1068-2341**

National Board Certified Teachers and Their Students' Achievement

Leslie G. Vandevoort
Audrey Amrein-Beardsley
David C. Berliner
Arizona State University

Citation: Vandevoort, L. G., Amrein-Beardsley, A. & Berliner, D. C. (2004, September 8). National board certified teachers and their students' achievement. *Education Policy Analysis Archives, 12*(46). Retrieved [date] from http://epaa.asu.edu/epaa/v12n46/.

Abstract

Contemporary research on teaching indicates that teachers are powerful contributors to students' academic achievement, though the set and interrelationships of characteristics that make for high-quality and effective teaching have yet to be satisfactorily determined. Nevertheless, on the basis of the extant research and a vision of exemplary teaching, the *National Board for Professional Teaching Standards* stipulated a definition of a superior teacher. The Board did this without empirical evidence to support their claim that teachers' who meet the standards set by the Board were superior in promoting academic achievement to those who did not meet those standards. In the 17 years since the founding of the National Board, only a few empirical studies have addressed this important issue. In this study we compare the academic performance of students in the elementary classrooms of 35 National Board Certified teachers and their non-certified peers, in 14 Arizona school districts. Board Certified teachers and their principals provide additional information about these teachers and their schools. Four years of results from the Stanford Achievement Tests in reading, mathematics and language arts, in grades three

through six, were analyzed. In the 48 comparisons (four grades, four years of data, three measures of academic performance), using gain scores adjusted for students' entering ability, the students in the classes of National Board Certified Teachers surpassed students in the classrooms of non-Board certified teachers in almost three-quarters of the comparisons. Almost one-third of these differences were statistically significant. In the cases where the students of non-Board certified teachers gained more in an academic year, none of the differences found were statistically significant. Effect size, translated into grade equivalents, informs us that the gains made by students of Board Certified teachers were over one month greater than the gains made by the students of non-Board certified peer teachers. Teachers identified through the assessments of the National Board for Professional Teaching Standards are, on average, more effective teachers in terms of academic achievement, one of the many outcomes of education for which teachers are responsible. This study does not address whether other, cheaper, or better alternatives to the National Boards exist, as some critics suggest. On the other hand, the results of this study provide support for the policies in many states that honor and provide extra remuneration for National Board Certified Teachers.

Introduction to the Policy Context and the Empirical Issue

In three short years America went from worrying about being *A Nation at Risk* (United States Commission on Excellence in Education, 1983) to worrying about how to become *A Nation Prepared* (Carnegie Forum on Education and the Economy, 1986). The first of these two reports focused attention on student academic achievement, purported to be too low for the economic viability of the United States, while the second report suggested that the best way to improve America's educational system was to focus on the quality of our nations' teaching force. Echoing both these themes about 20 years later, the No Child Left Behind Act (NCLB), enacted by Congress in 2002, challenged Americans to put a "highly qualified teacher" in every classroom by the year 2006.

Before and after the reports of the 1980's, the community of researchers concerned with teaching produced many studies to determine the relationship between teacher variables and student achievement. Hanushek (1992), for example, estimated that a high quality teacher, in comparison to a low quality teacher, can provide one full years difference in the learning of a class of children (one and one-half years growth in grade level vs. only a half years growth). Others echoed this theme (e. g. Goldhaber, 2002; Ferguson, 1998). While no single approved list of characteristics has emerged, it is generally agreed that credentials alone (graduation from a particular school of education, having advanced course work in education, holding a masters of education degree) do not provide assurance about the qualifications of teachers. Other factors are at work (Goldhaber and Brewer, 1996; 2000). But in the end, wrote Katie Haycock for the Education Trust (1998), "…What all of the studies conclude, is the single most important factor in student achievement (is) the teacher." (p. 2). This claim has many supporters.

The National Board for Professional Teaching Standards. The National Board for Professional Teaching Standards (NBPTS) grew out of the emergent belief that teachers were a key factor in improving student achievement, and thus the profession needed ways to recognize and appropriately reward exemplary classroom teachers. The NBPTS was created in 1987 at the

recommendation of the Carnegie Task Force on Teaching as a Profession. One of its primary goals was to "identify and recognize teachers who effectively enhance student learning" (http://www.nbpts.org/events/qabrochure.cfm). Teachers who hold certification from the Board are expected to have demonstrated "the high level of knowledge, skills, abilities and commitments" (p. 2) that are reflected in the Board's five core propositions. National Board Certified teachers (NBCTs) are teachers who: 1) are committed to students and their learning; 2) know the subjects they teach and how to teach those subjects to students; 3) are responsible for managing and monitoring student learning; 4) think systematically about their practice and learn from experience; and 5) are members of learning communities. Three of these propositions directly relate to student achievement. Strongly implied, but not stated explicitly, is that NBCTs are superior teachers.

The NBPTS maintains that only those teachers who have proven their ability to enhance student learning earn Board certification status. The expectation, therefore, is that the students of Board-certified teachers will make yearly achievement gains that tend to be greater than those obtained by the students of teachers who have not undergone the demanding Board certification process. This is the empirical issue addressed by this study. But it ought to be obvious that the NBPTS does not need student achievement data to claim it is identifying exemplary teachers. Medical boards, for example, inform us only that a physicians passing score has the potential of providing high quality care in some specialty area. The health of the patients of medical board certified physicians is rarely assessed. Board certification in medicine and in education is defensible without student outcome data. Nevertheless, we decided to focus on student outcome data for NBCTs, providing a more rigorous test of board certification in education than is typically done in medicine, law or accountancy.

A Brief History of the National Board for Professional Teaching Standards. In the early 1980's the United States was suffering from one of the worst economic recessions since the Great Depression. Interest rates were at 21%, inflation was in the double digits, and the unemployment rate exceeded ten percent. By 1983 approximately one third of America's industrial capacity remained idle. At the same time, however, the total number of computers in the US increased from one million units to over ten million. New technology resulted in worldwide increases in communication and information. Political systems were becoming increasingly more democratic and a global economy was developing rapidly.

In order to remain a viable player in the world market it was thought that Americans needed to know more than the basic reading, writing and mathematics skills characteristic of an industrial era. It was predicted that workers of the 21st century would need to formulate new information from that which already existed, be problem-solvers, and engage in collaborative activities. Many believed that the United States public education system would need to make major changes in order to accomplish the task of preparing students for the work demanded in the next century. Hence, in 1985, the Carnegie Forum on Education and the Economy established the "Task Force on Teaching as a Profession." Members of the task force came from teacher organizations and unions, other educational organizations, as well as government and business. The primary goals of the task force were to remind Americans of the economic challenges that lay before them, reinforce education as fundamental to the growth of the economy, reaffirm the teaching profession as "the best hope for establishing new standards of excellence…" (p. 7), and make Americans aware of the opportunity to reform education, a chance that might not present itself again until well into the 21st century.

The task force report, entitled *A Nation Prepared: Teachers for the 21st Century* (Carnegie Forum on Education and the Economy, 1986), was a monumental plan for restructuring schools and the teaching force. The task force foresaw "an economy based on people who think for a living." The school would become the place where students developed their thirst for knowledge. Teachers

would be critical to this learning process and as such, they would be expected to think for themselves, act independently as well as collaboratively, possess a knowledge base of both depth and breadth, be able to communicate their knowledge, stimulate others to achieve, and be able to think and act with critical judgment (p. 25). Teachers of the future would be expected to create learners who could take newly acquired knowledge and apply it to novel problems and situations. Students would no longer be passive learners and teachers merely the purveyors of information. The task force portrayed the teacher as an agent of social change.

A Nation Prepared outlined a plan of action, including the establishment of a National Board for Professional Teaching Standards. The Board was to consist primarily of teachers, but would include others from state and local education agencies as well as from business and higher education. Early supporters of the plan included Mary Hatwood Futrell, then-president of the National Education Association; Albert Shanker, then-president of the American Federation of Teachers; and Lee Shulman of Stanford University, who was then conducting research on novel forms of teacher assessment. In addition to these individuals was one of the Board's strongest supporters, North Carolina's Governor Jim Hunt. It was he who was instrumental in seeking funding for the Board from the Carnegie Corporation of New York (see National Board for Professional Teaching Standards, no date).

Once funding was in place, a 33 member planning team began the process of writing bylaws and articles of incorporation. Additional members were added to the group and the Board held its first meeting in October 1987. Its vision was sweeping. It sought to improve the public's perception of teachers, restore faith in public education and instill in teachers an improved sense of self-esteem. It would do so by increasing the knowledge base for teaching and by encouraging the development of more rigorous teacher education and professional development programs. The Board also envisioned an increase in the numbers of top-quality individuals entering the teaching field as well as a decrease in the numbers of excellent teachers leaving the field. At the center of the Board's vision and high on its agenda was the concept of a national teacher certification system. The model the Board had in mind was that of the medical professions' National Board examinations for specialty areas such as oncology, hematology, family practice, and so forth.

In a document prepared by the Board entitled, *Towards High and Rigorous Standards for the Teaching Profession* (National Board for Professional Teaching Standards, 1989), it was stated: "While many noteworthy efforts are being made to improve schools, none promises the potential for permanent and systemic transformation of teaching that is offered by the National Board: to establish high and rigorous standards for what teachers should know and be able to do and to certify teachers who meet those standards" (p.iii). The NBPTS standards now define accomplished teaching in 27 different specialty fields and these activities have been described as the "heart of the work" of the Board.

With assistance from researchers in teaching, members of the Board began debating the concept of what constituted an accomplished teacher. That task, as well as the development of the Board's standards and assessments, took over six years to complete and involved extensive time commitments by expert teachers, school administrators and scholars. The standards were initially presented as drafts that were reviewed by individuals within education, members of the non-teaching community, and members of the NBPTS Board of Directors. In the final analysis it was estimated that each of the final standards documents in a specialty area, and the accompanying assessment instruments, involved the work of hundreds of individuals, at the costs of millions of dollars.

Potential candidates for Board certification would need to familiarize themselves with the standards in their area of teaching, put them into practice within their classrooms, and then begin the Board's assessment process leading to certification. The process would be voluntary, and teachers who sought the certificate would be expected to engage in intense self-reflection and

analysis of their own teaching. They would also be expected to demonstrate their ability to enhance student learning by adhering to the Board's five core propositions. As proposed in 1989, these propositions are:

1. Teachers are committed to students and their learning. Accomplished teachers are dedicated to making knowledge accessible to all students. They act on the belief that all students can learn. They treat students equitably, recognizing the individual differences that distinguish one student from another and taking account of these differences in their practice. They adjust their practice based on observation and knowledge of their students' interests, abilities, skills, knowledge, family circumstances, and peer relationships.

Accomplished teachers understand how students develop and learn. They incorporate the prevailing theories of cognition and intelligence in their practice. They are aware of the influence of context and culture on behavior. They develop students' cognitive capacity and their respect for learning. Equally important, they foster students' self-esteem, motivation, character, civic responsibility, and their respect for individual, cultural, religious and racial differences.

2. Teachers know the subjects they teach and how to teach those subjects to students. Accomplished teachers have a rich understanding of the subject(s) they teach and appreciate how knowledge in their subject is created, organized, linked to other disciplines and applied to real-world settings. While faithfully representing the collective wisdom of our culture and upholding the value of disciplinary knowledge, they also develop the critical and analytical capacities of their students.

Accomplished teachers command specialized knowledge of how to convey and reveal subject matter to students. They are aware of the preconceptions and background knowledge that students typically bring to each subject and of strategies and instructional materials that can be of assistance. They understand where difficulties are likely to arise and modify their practice accordingly. Their instructional repertoire allows them to create multiple paths to the subjects they teach, and they are adept at teaching students how to pose and solve their own problems.

3. Teachers are responsible for managing and monitoring student learning. Accomplished teachers create, enrich, maintain, and alter instructional settings to capture and sustain the interest of their students and to make the most effective use of time. They also are adept at engaging students and adults to assist their teaching and at enlisting their colleagues' knowledge and expertise to complement their own. Accomplished teachers command a range of generic instructional techniques, know when each is appropriate, and can implement them as needed. They are as aware of ineffectual or damaging practice as they are devoted to elegant practice.

They know how to engage groups of students to ensure a disciplined learning environment, and how to organize instruction to allow the schools' goals for students to be met. They are adept at setting norms for social interaction among students and between students and teachers. They understand how to motivate students to learn and how to maintain their interest even in the face of temporary failure.

Accomplished teachers can assess the progress of individual students as well as that of the class as a whole. They employ multiple methods for measuring student growth and understanding and can clearly explain student performance to parents.

4. Teachers think systematically about their practice and learn from experience. Accomplished teachers are models of educated persons, exemplifying the virtues they seek to inspire in students -- curiosity, tolerance, honesty, fairness, respect for

> diversity and appreciation of cultural differences -- and the capacities that are prerequisites for intellectual growth: the ability to reason and take multiple perspectives to be creative and take risks, and to adopt an experimental and problem-solving orientation.
>
> Accomplished teachers draw on their knowledge of human development, subject matter and instruction, and their understanding of their students to make principled judgments about sound practice. Their decisions are not only grounded in the literature, but also in their experience. They engage in lifelong learning that they seek to encourage in their students.
>
> Striving to strengthen their teaching, accomplished teachers critically examine their practice, seek to expand their repertoire, deepen their knowledge, sharpen their judgment and adapt their teaching to new findings, ideas and theories.
>
> **5. Teachers are members of learning communities.** Accomplished teachers contribute to the effectiveness of the school by working collaboratively with other professionals on instructional policy, curriculum development, and staff development. They can evaluate school progress and the allocation of school resources in light of their understanding of state and local educational objectives. They are knowledgeable about specialized school and community resources that can be engaged for their students' benefit, and are skilled at employing such resources as needed.
>
> Accomplished teachers find ways to work collaboratively and creatively with parents, engaging them productively in the work of the school.

According to the Board, these five propositions "articulate what teachers should know and be able to do" and are an "expression of ideals" that guided the development of both the standards and the assessments in each specialty area.

Assessments for Certification. The Board's assessment process is performance-based and includes the evaluation of portfolio entries as well as the completion of a set of tasks that take place at an assessment center, usually over the course of a full day. As part of the portfolio assessment, teachers videotape and analyze their teaching, provide evidence of student learning, and display artifacts used in their teaching. The portfolio portion of the assessment was designed to examine the ways in which teachers put theory into practice in their classrooms.

Testing at the assessment center requires teachers to provide written responses to questions that are specific to their field of teaching. The Board's goal in developing these activities was not only to complement and expand upon the portfolio, but also to allow the candidates the opportunity to demonstrate the scope of their content-specific knowledge.

Performance tests such as those chosen by the board are expensive to develop and to score. Thus, for teachers, the costs to take the examination are high, currently running about $2,300.00. To successfully complete the certification process, the candidate is required to earn a minimum score on all of the sub-sections of the portfolio assessment and on various sub-tests taken at the assessment center.

The first group of teachers to obtain National Board certification status did so in 1994. There were less than 100 teachers certified in that first year (Goldhaber & Anthony, 2004). Each year the numbers have increased and in the most recent year, 2003, over eight thousand teachers earned certificates in the 27 different specialty areas that are tested by the NBPTS. The total number of NBCTs in the nation is now about 32,000 out of about 65,000 teachers who took the exam. The first-time passing rate for the Boards is around 48 percent (Goldhaber & Anthony, 2004).

Research on National Board Certified Teachers

The vast majority of reports on NBCTs and those that compare them with non-NBCTs are favorable. But most have the problem of being self-reports from individuals who went through a rigorous assessment process and passed. These "winners" are more likely to feel persuaded that their experience has helped them to be better teachers. Thus, many of these studies are of little significance in understanding characteristics of the NBCTs in comparison to their non-Board certified peers. On the other hand, although not at all conclusive, the consistency of the data across researchers and research methods provides reviewers more confidence that the data obtained from these studies is worth taking seriously. Even those who take the Boards and fail report that the process is a valuable learning experience. None of the studies in this section of the review, however, address the question that always looms large for the NBPTS, namely, what are the effects of NBCTs on their students' achievement? We will report the few studies that exist about this issue later in this review.

Research without student outcome data. Examples of the kinds of research completed where outcome data for students were not reported include the Boards own inquiry, in 2001, in which nearly 5,000 teachers who had achieved certification prior to the year 2000 were surveyed. Among other things, these teachers responded to questions regarding their involvement in leadership activities. Ninety-nine percent of the teachers reported that since their certification they had become involved in at least one leadership role aimed at improving teacher quality or student achievement. Eighty-nine percent of the teachers agreed that increased involvement in leadership activities made them a more effective educator. Sato, Hyler and Monte-Sano (2002) also document how the Board process and status as a NBCT enhances a desire to lead and the opportunities to lead. Data from these studies suggest that others in the educational community regard NBCTs as qualified to lead, an indirect way of affirming the competence of the NBCTs.

In an attempt to determine attitudes to teaching and other related activities, Whitman (2002) surveyed nearly 2000 teachers, both NBCTs and non-NBCTs. She found that NBCTs were less likely to believe that external conditions determined the educational outcome of schooling. Similarly, she also found that NBCTs were more likely to believe that each child could be taught successfully, regardless of their home situation or other external factors. Their perceived self-efficacy was high. Whitman concluded that NBCTs were accomplished individuals with a greater sense of responsibility for their classrooms, greater commitment to their careers, and greater professionalism and collegiality than the non-NBCTs. Her study suggested that districts benefit from having NBCTs in their employ.

In the state of Florida, Ralph (2003) surveyed 239 NBCTs in an effort to determine their views of the Board certification process and its effect on the professional culture of the schools. He found that a majority of NBCTs viewed all of the Board certification activities as "very" or "somewhat" important. The NBCTs also indicated that they "almost always" or "frequently" experienced the specifically identified elements of a professional culture as they went through the certification process. He also reported that NBCTs had greater needs for leadership activities than the their non-NBCT peers.

The Indiana Professional Standards Board (2002) surveyed NBCTs and found that they believed that the certification process had made them more effective teachers. One fourth of these teachers reported liking challenges and nearly this many, 22%, considered themselves to be lifelong learners. A majority of the NBCTs (62.5%) reported the greatest benefit of Board certification to be the increase in the number of professional opportunities that were made available to them.

With an estimated one-third of its teachers leaving the profession within the first 5 years, there is a critical need to determine ways in which to retain good teachers in the state of North

Carolina. This was one of the goals of a group of studies carried out at the University of North Carolina (UNC) and funded by a grant from the Board. Petty (2002) surveyed two groups of North Carolina's high school math teachers, those with Board certification and those without. When asked about their wants and needs for satisfaction and success in their career, Petty found several similarities between the groups. Both groups mentioned the need for administrative support to obtain adequate materials, salary increases, smaller class sizes, etc. Both groups also had similar professional development needs. The NBCTs, however, differed from their non-NBCT peers in that they sought recognition for their achievement, perceived their work as being an integral part of their lives, and desired leadership roles in professional development activities.

Sampling North Carolina's third to fifth grade NBCTs and non-NBCTs, O'Connor (2003) found differences in their levels of education. The NBCTs were more likely to have attained masters' or doctorates than their non-NBCT peers. She also found that more NBCTs had earned teaching honors, particularly the "School Teacher of the Year" award, than non-NBCTs. To foreshadow an issue that arises later, it may be that NBCTs are already quite able before they engage in a year-long process to become certified.

Other findings from this study indicated that NBCTs viewed autonomy as a more important need than non-NBCTs. The NBCTs also reported the need for time for individual study more frequently than non-NBCTs and NBCTs were more likely to read professional journals than their peers.

The third researcher in this set, Dagenhart (2003), found a relationship between Board certification and the need for autonomy in middle school teachers. She too found NBCTs reporting that they have a greater need to read professional journals than non-NBCTs. All three researchers from UNC reported that NBCTs were more likely to seek opportunities for leadership than non-NBCTs.

Other researchers have used case studies, observation, interviews, and other methods to gain further information about Board-certified teachers. For example, Turchi (1996) found that teachers undergoing the certification process developed new insights into their professional relationships. They also became more aware of the administrative power within their school and in educational institutions in general.

In 1998, Iovacchini studied nine NBCTs to determine what they had learned from the certification process and how the process had contributed to their professional growth. She found that teachers who earned Board certification had made changes in their teaching that reflected the Board's standards. The NBCTs typically went through the process for intrinsic reasons, developed a stronger sense of confidence in their teaching, and developed new understandings about curriculum. They reported having made gains in pedagogical knowledge and in their collaborative activities with peers. The certification process, they claimed, allowed teachers to have a broader perspective of their roles and of their needs. Iovacchini believed that this small sample of teachers were committed professionals who used the standards as benchmarks for their own performance.

Multiple case studies were used by Taylor (2000) to examine teacher change in 11 NBCTs from Colorado. She found that through reflection, many of the teachers reexamined their previous teaching practices. Taylor did not examine teachers' views on standardized tests but did confirm their use of various other measurement strategies, including prompts, work samples, journals, drama, story telling and rubrics. All of the NBCTs in this study reported having made changes in the ways they assessed students after having gone through the Board certification process.

Some moving testimony about how the Board certification process changed teachers and teaching is found in Sato (2000). These constructive transformations are also described by Buday and Kelly (1996), while survey data showing overwhelming support for the Board process has been

collected by Belden (2002). Ways to enhance the Board process are described by Berg (2003). In sum, these studies make a convincing case for the positive effects of the Board certification process.

But the process is not without problems. Using a case study design, Burroughs, Schwartz and Hendricks-Lee (2000), followed four candidates for Board certification to determine how they perceived their certification tasks and what problems resulted from those tasks. They interpreted the difficulties the candidates encountered in the certification process as resulting from the need to move from a world of practice (teaching) to a world that valued discourse (writing). Findings indicated that candidates had difficulty portraying their teaching in written form and that the discourse required by the Board may have been at odds with teachers' practical knowledge.

Teachers' practical knowledge is often described as situated, interpersonal, and tacit (Van Driel, Beijaard, & Verloop, 2001), thus it is often hard for teachers to verbalize (or researchers to determine) what teachers actually know and are able to do. In the Burroughs et al., study, those who were best able to overcome the problem of articulating their knowledge were those who were best able to accept the Board's discourse values. Research like this suggests a greater-than-usual likelihood that false negatives and false positives will be identified through the Board certification process. Some highly articulate teachers may become certified and some less articulate teachers may miss out, and the fault may be in their ease or difficulty in translating personal, practical, knowledge-in-action into a form of knowledge-about-action that is amenable to assessment in conventional ways (see Schön, 1983).

Pool, Ellet, Schiavone and Carey-Lewis (2000) interviewed a small sample of NBCTs to determine if the manner in which the Board certified its candidates, via portfolios, center exercises, etc., was a valid way in which to measure the behaviors or actions of teaching. The authors were particularly interested in examining differences among the NBCTs. Their method of study involved teacher observations and interviews with the colleagues of the NBCT. They found considerable variability between teachers in regards to the quality of teaching and learning that took place in their classrooms. In fact, when applying Berliner's descriptors of expert teachers (1994a), they concluded that this group of NBCTs ranged from novice to expert in skill level. This research study, as in that of Burroughs et al. (2000) described above, suggests the possibility of a greater-than-usual number of false positives being among the NBCTS. Pool et al. believed, however, that those NBCTs who most valued the NBPTS philosophy were those who practiced a higher level of teaching and maintained a higher-quality learning environment than did those who could not articulate the Board's values as well.

The evidence that NBCTs have some exemplary characteristics is strong, although it is likely they may have been that way before they took the Boards. Nevertheless, the Board process--the preparation for taking the Board exams—appears to have independently made contributions to their practice. Revealed also in this literature is the possibility that the certification examination may identify more than the usual numbers of false negatives and false positives, an issue that we think deserves much more research. There is a growing belief, based on recent research (Shutz and Moss, 2004; Moss et al, 2004), that the kinds of assessments used by the NBPTS gives only a brief glimpse of what a teacher is capable of under restrictions and controls. Typical, everyday classroom performance must necessarily differ from the performance displayed and judged from portfolio's and at assessment centers. It is not that the two are unrelated, but rather that classrooms are very complex settings, subject to many day-to-day changes and they are particularly variable from year to year or school period to school period. It should not be surprising, therefore, that a relatively inexpensive and relatively short assessment yields false negatives and positives. There are ways to decrease the numbers of such misclassifications but they typically require more time and money for assessment.

Research examining student outcomes. In contrast to the surveys, interviews, and case studies noted above, are a small number of research studies that address the "outcomes" question, the question that always appears to be looming in the background and influencing the acceptance of the entire Board certification process. It is of course appropriate to point out that neither the Medical Boards, Bar Examination, nor the examinations to obtain certification as an accountant, real estate broker, or cosmetologist address this question. It is the teachers' examination, more than any of the others, that is held to this most rigorous and possibly unreasonable standard. Regardless of its problematic nature, the demand exists for assurances that NBCTs' have positive effects on their students' academic outcomes.

One of these outcome studies was done by Stephens (2003), in which she assessed the relationship between Board certification and math achievement as measured by South Carolina's Palmetto Achievement Challenge Test (PACT). Using an ex-post facto design, Stephens analyzed test scores from fourth and fifth grade students in two school districts in the state. She compared the 2002 PACT scores from students of NBCTs to those of non-NBCTs, first matching students on their 2001 PACT scores. She attempted to control for the teachers' level of experience as well as the poverty level of the schools involved in the study. The scores of 154 students of NBCTs were compared to the scores of 669 students of non-NBCTs and in 87 percent of the comparisons, there was no significant difference between the achievement of the two groups of students. The study, however, was not sophisticated statistically; the researcher seemed to have had difficulty in matching; and in some of the comparisons between NBCTs and non-NBCTs the numbers of students for which there were data were quite small, and so statistical power was severely limited.

The study that has probably drawn the greatest amount of attention was that done by Bond and his colleagues (Bond, Jaeger, Smith and Hattie, 2000; Bond, Smith, Baker and Hattie, 2000). In our judgment this is a unique and creative study, having as its major flaw that it was funded by the NBPTS and that the lead authors worked with the Board over many years. This allows critics (see below) to question the objectivity of the study. (In the interest of full disclosure, it is noted that the third author on this study was a consultant to the Board for two years when it began its assessment program. In addition, he served also as a consultant to the designers of the particular study reviewed here because the literature on teacher expertise provided a basis for the design of the study. Finally, The Boards' small grant program provided some of the funds for this study, though none of the three authors had contact with the board on any substantive matter from the day of funding until the day the study was completed.)

On the basis of the literature on expertise and on teachers' content and pedagogical content knowledge (Berliner, 1994a; b; Shulman, 1987; Shulman and Quinlan, 1995), Bond and his colleagues choose to specify expert classroom performance as consisting of a number of prototypic characteristics. Bond and his colleagues invented unique measures to assess each prototypic feature of expert teachers. For example, following the logic of Sternberg and Horvath (1995), among others, Bond and colleagues asserted that the expert teacher (like other experts) has extensive and accessible knowledge. For teachers this would be knowledge about classrooms, subject matter, and classroom context. Trained observers and analysts assessed this feature by analyzing and numerically coding teachers' classroom lessons and transcripts that were obtained from interviews with the teachers in this study. In this case highly trained raters were looking for evidence of organization and re-organization of knowledge, connections of the teachers' knowledge to other school subjects, and the connection of the teachers knowledge to the prior and future learning of their students.

A total of thirteen prototypical features of expertise were hypothesized, and measures were created for each feature. For each prototypical feature raters were trained to acceptable

levels of reliability and performed their analyses blind with regards to the skill level of the teachers they were assessing. The thirteen prototypic features hypothesized to be held by expert teachers were:

- better use of knowledge,
- extensive pedagogical content knowledge, Including deep representations of subject matter knowledge,
- better problem solving strategies,
- better adaptation and modification of goals for diverse learners, better skills for improvisation,
- better decision making,
- more challenging objectives,
- better classroom climate,
- better perception of classroom events, better ability to read the cues from students,
- greater sensitivity to context,
- better monitoring of learning and providing feedback to students,
- more frequent testing of hypotheses,
- greater respect for students, and
- display of more passion for teaching.

The outcomes of instruction for students of expert teachers were hypothesized as well. These included:

- higher motivation to learn and higher feelings of self-efficacy,
- deeper, rather than surface understanding of the subject matter, and
- higher levels of achievement.

To assess the occurrence of these prototypic features of expert teachers, two samples of teachers were recruited from among those who had attempted to obtain National Board Certification in the areas of Middle Grade Level/Generalist, or Early Adolescent Level/English Language Arts. One of the comparison groups (N= 31) consisted of those who passed the National Board assessments; the other comparison group consisted of those who did not achieve Board certification through the assessments (N=34). All the teachers were well experienced, had prepared diligently for the examinations, and spent considerable amounts of time and money to demonstrate they were highly accomplished teachers. This is important because the comparisons of the occurrence of prototypical features of expertise, and of the student outcomes of the two groups, were not between expert and non-expert teachers. These comparisons were between equally experienced, well-prepared teachers, all of whom thought they were highly accomplished. Thus, this was a very conservative investigation of whether the Board assessments could really identify expertise in teaching.

We believe that the results of this study are quite remarkable. The Board certified teachers, in comparison to those that failed to meet the Board standards on the assessments, excelled on each and every prototypical feature, with statistical significance found in 11 of the 13 comparisons of the features. When looked at as effect sizes, the differences between these two highly experienced groups ranged from just over one-quarter of a standard deviation to 1.13 standard deviations in favor of the Board certified teachers. Thus, teachers found to be expert on the basis of the assessments used by the NBPTS were anywhere from 8 percentile ranks to 37 percentile ranks higher on measures of their use of knowledge, the depth of their representations of knowledge, their expressed passion, their problem-solving skills, and so forth.

When a discriminant function was used, it was found that about 85 percent of the highly experienced, well-prepared teachers comprising these two groups could correctly be discriminated from each other. The features with the greatest ability to discriminate between the NBCTs and those that had failed to earn certification were the degree of challenge that the curriculum offered, the teachers' ability for deep representations of the subject matter, and the teachers' skillfulness in monitoring and providing feedback to his/her students. This study provides validity for the assessment program but, as described below, the study fell short when it comes to documenting the NBCTs effects on student outcomes.

Over a dozen scales were used to measure the motivation and self-efficacy of the students of these two groups of teachers. On these important student outcomes of schooling few differences between those who passed certification and those who did not appeared.

Student academic achievement was evaluated through written assignments. But covariates reflecting initial ability of the students could not be obtained, so these data are not completely convincing. Nevertheless, the Board certified teachers had students who performed better on the writing assignment although the mean scores for the two groups did not differ significantly.

On the analysis of student work samples, however, 74 percent of those obtained from the students of NBCTs demonstrated higher understanding through more relational and more abstract student work. Only 29 percent of the work samples from the students of the non-NBCTs showed these characteristics. Bond and his colleagues expressed their belief that when compared to standardized test scores, "surface versus deep" understanding of objectives is an equally defensible dependent variable. Bond, Smith, Baker and Hattie (2000) concluded from these data that the NBPTS, through its assessments, is

> identifying and certifying teachers that are producing students who differ in profound and important ways from those taught by less proficient teachers. These students appear to exhibit an understanding of concepts targeted in instruction that is more integrated, more coherent, and at a higher level of abstraction than understanding achieved by other students. (p. 113).

While we agree with this conclusion, it is also true we have no knowledge about the demographic characteristics and the academic achievement of the students that were in the classes of those teachers. This is a design problem that suggests caution in interpreting the findings. However that design problem is likely not to influence the other major conclusion of the study, namely, that teachers designated as experts from assessments developed by the NBPTS met the criteria for expertise set forth in the prototypic model.

A second major study supporting the NBPTS used data available from North Carolina's Department of Public Instruction. Goldhaber and Anthony (2004) studied the relationship between Board certification status and student achievement at the elementary school level. The data set consisted of grade 3-5 teachers' administrative records, as well as about two hundred thousand elementary student test scores for each of three years (1996-1997 to 1998-1999). The total pre- and post-test merged teacher and student records in the areas of reading and mathematics were each over 600,000. The researchers examined the value added by NBCTs versus non-NBCTs, including those who had attempted certification and failed and those who had never applied.

For the three years in which data were examined, the authors found that students of NBCTs significantly outperformed those of their non-NBCT counterparts, whether they had taken the test or not. This finding led Goldhaber and Anthony to conclude that in regards to student achievement gains, NBCTs were more effective than their peers.

Advantages accruing to the students of NBCTs on the state test, compared to the students of other teachers in the state, were modest, but consistent. But these data are complex, revealing interactions that complicate interpretation. The researchers found that the significance and magnitude of the "Board" effect varied according to grade level and student characteristics. For example, younger and poorer students profited more from NBCTs than did older and wealthier children. We hypothesize that the more teaching skills that are needed (for example, in the teaching of younger and poorer students), the more NBCTs seem to excel their peers.

Another finding of interest was that current NBCTs were no more effective than non-NBCTs during the year that board certification was undertaken. Perhaps the time and effort that goes into Board certification reduces the future NBCTs effectiveness for that year. This issue certainly needs further research.

This study also sheds light upon another issue of interest, namely, whether NBCTs are already more effective teachers or become more effective by engaging in the Board assessment process. From these data it appears that the NBPTS is identifying teachers who already were more effective in the state of North Carolina. It is also worth noting that Goldhaber and Anthony tested models that addressed the hypothesis that NBCTs have higher achieving students to teach, and work in schools that are more advantaged. While there was some evidence that this was true, the researchers also found that these factors were statistically unrelated to the gains made by the students of NBCTs.

These researchers believed that their investigation used rigorous methods and found robust enough results so that the controversy regarding national certification and its relationship to student achievement could be put to rest. The researchers believe that their findings confirm that the NBPTS was, indeed, identifying and certifying teachers who raise student achievement.

Criticisms of the Board's assessments. The study by Bond, Smith Baker and Hattie (2000) produced a flurry of comments, most notably one by Michael Podgursky (2001), entitled "Defrocking the National Board." Podgursky, professor of economics at the University of Missouri-Columbia, was highly critical of the methodology used in Bond's study. He considered it a "dubious proposition" for the authors to refer to the meta-analysis of 200,000 studies as providing scientific evidence of the 13 dimensions of teaching expertise. Podgursky referred to the measures of student performance as "vague" and discussed possible sources of bias in the study due to lack of controls for the students' previous achievement level, socio-economic status or demographic characteristics. He also maintained that the two groups being studied, the NBCTs and those who had attempted certification but had not succeeded, were not randomly chosen but were selected in such a way as to increase the chances of finding an effect due to certification. Podgursky also cited the scarcity of independent research on the Board as well as large investments made by states to reward teachers for becoming Board certified as additional areas of concern. In conclusion he called for a "rigorous and arm's length cost-benefit study of National Board certification" (p. 8).

Six months later (and well before the Goldhaber and Anthony study, noted above) Podgursky (2001) published another article entitled, "Should States Subsidize National Certification?" Again, he found fault with the Board and the certification process, a process he believed to have resulted from the teacher unions' dissatisfaction with merit pay. He attacked the Board for awarding certificates to teachers who had not demonstrated significant knowledge of content or whose writing contained errors of syntax or grammar. He found fault with the Board for ignoring input from parents and principals, saying it allowed candidates "extensive opportunities for cheating." Podgursky also bemoaned the fact that the Board had developed national education standards, ignoring the standards already developed by individual states. In answer to his initial question about state subsidization of certification, he stated,

> In spite of the fact that the board has been in operation since 1987, and has received nearly $100 million in federal support, no rigorous study has ever been undertaken to determine whether the students of board-certified teachers actually learn more than students of an average teacher in the workforce (or teachers who have failed the board assessment), where student achievement is measured by a state assessment or a standardized objective exam. Nor do we have evidence that this costly and time-consuming process is actually any better at identifying superior teachers than assessments from supervisors, principals, or parents (p. 3).

Bond (2001) replied to Podurgsky's claims. He provided 30 pages of detailed descriptions of the 13 teaching dimensions as evidence that the dimensions were not "vague," but were based on competent educational research. The fact that the dimensions appeared to be a repetition of the Board's certification process was further proof to Bond that the Board had developed a valid method of assessing accomplished teaching. Bond went on to address Podgursky's claim that issues of SES and student achievement were not adequately addressed and we think he refuted these adequately. In the end, Bond made "no apologies whatsoever" (p. 5) when commenting on the quality of these procedures used in the study.

Also responding to Podgursky was Betty Castor (2001), then-president of the Board. She maintained that Podgursky was unaware of the facts and disputed seven of Podgursky's claims. Specifically Castor provided evidence of the market-driven nature of the certification process, the intent of the process to strengthen the teaching profession, and to help create excellent teachers. She provided evidence of the specificity and rigor of the Board's standards and disputed the claim that the standards were intended to supercede those of any individual state. Castor described the Board's assessment procedures as "controlled and secure" (p. 3), and its evaluations as being objective and measurable. In conclusion, Castor asserted that National Board Certification was a process that was proving to have a positive impact on student learning (Although the data for that claim, at that time, was more circumstantial and indirect. Castor's claim is now better warranted given both the Goldhaber and Anthony study, described above, and the present study).

In 1998 Dale Ballou wrote for *Education Week,* noting the increasing resources and influence of the Board. He maintained, however, that the Board had not been able to provide answers to a number of questions, including whether or not it was actually able to identify superior teachers. In the article he referred to the Board's standards as "vague" (p.1) and went on to discuss the subjective nature of performance assessment in general. He questioned the Board's citation of numerous validity studies as "based entirely on the opinions of panels of educators as to what an accomplished teacher should know and do, not on objective measures of student performance" (p.2). Ballou appears not to understand how other professions develop their certification procedures. For example, law, medicine, and accounting have little reliable evidence linking their testing procedures to outcomes as an attorney, physician or accountant. Nevertheless, criticisms of this type continue to foster an interest in studying student outcome data.

Ballou also questioned whether Board certified teachers were better than the average teacher simply because they came from a superior group of applicants. Thus, he asks, was the certification process responsible for creating better teachers or were they better than average teachers before they became certified? This question is interesting, but also appears to us to be irrelevant, given the NBPTS' goal of identifying exemplary teachers. The process leading up to Board certification may be extremely informative to teachers, as the vast majority of successful (and unsuccessful) candidates taking the assessments affirm. But that can also be considered merely a beneficial side effect. The major issue, we think, is simply whether the Board does in fact certify exemplary teachers, not

whether the assessments are instructive or whether the successful candidates were exemplary before engaging in preparation for the Boards.

While the primary goal of the assessment seems to us to be identification of exemplary teachers, it is also true that the process of preparing for the assessments should be an opportunity for high-quality professional development. At least one critic thinks that the Boards have lost an opportunity to do that job as well as it could have. Petrosky (1998) was Principal Investigator and Co-Director of the Early Adolescence English Language Arts (EAELA) Assessment Development Lab for NBPTS. He played a primary role in developing the first certification tests for English teachers. His work was based on the Board's initial vision of the certification process as being a significant professional growth activity for teachers. As his work progressed he came became more convinced that peer assessment would lead to a long term change in the culture of education and as such, would bring about greater reform. He was, therefore, critical of the Board's 1994 decision to use the Educational Testing Service (ETS) as the only agency to develop and score Board certification assessments. He referred to ETS as "outsiders," professional test designers whose main concern was to sort and rank candidates with little regard for feedback of their performance. As a result of changing from teacher evaluators or "insiders," to professional scorers, a chance was lost for teachers to have ongoing interactions and discussions with their peers, as well as opportunities to rethink, review and revise the Board's standards. Petrosky argued that when the assessment process changed from the control of the teachers to the control of the assessment professionals, the certification process ceased to be a professional growth experience for all those involved, no matter what their role in the process.

Petrosky (1994) addressed these same concerns when he spoke at a conference of The National Council of Teachers of English. He reiterated his stance that the Board's original vision of the certification process was one of professional growth but went on to say that in expectation of thousands of teacher applicants, Board members decided to seek out a method of judging teachers that could be used on a large scale basis and would prove to be cost effective and relatively simple to administer. As a result of policy decisions such as this, Petrosky believed that Board "had reduced complex (teaching) performances to numerical ratings based on generalized rubrics with canned feedback." He offered alternative assessment strategies such as an adjudication process whereby one judge is responsible for all of the evidence collected regarding a number of candidates. That judge would review all of the teacher performance ratings from each of the other judges and then through the process of adjudication, determine a candidate's status. Another suggestion was to have a pair or panel of judges responsible for evaluating the total set of a candidate's assessments. Petrosky's final argument involved the policy makers. He believed that the Board needed to be staffed by teachers who could make executive decisions based on their vision, a vision that he believed could not be adequately represented by "bureaucrats and test makers."

J. E. Stone (2002), using a value-added method of defining successful teaching, asserted that no studies had proven that Board-certified teachers were able to improve student achievement in objectively measurable ways. In an effort to explore this issue, Stone examined "teacher-effect scores" for 16 NBCTs in Tennessee. Teacher-effect scores are estimates of the impact a teacher has on a student's learning. In Stone's words, "student progress is estimated on the basis of how much students gain in comparison to their achievement increases in previous years" (p. 1). For each of the 16 NBCTs, Stone reported the teacher-effect score, the standard error of measurement and the resulting percentage of annual achievement growth. In this study annual achievement growth was referred to as the "critical indicator of teacher effectiveness." Growth was calculated by determining the ratio of the teacher-effect score to the average annual achievement growth for the school district in which the teacher worked. That number was multiplied by 100. A score of 115, therefore, would indicate that the teachers' students had exhibited gains of 115% of the local mean. Teachers with

scores of 115 or more were considered "exemplary." Stone used this same criteria when examining the gains made by students of NBCTs. Of the 123 teacher-by-subject-by-year teacher-effect scores he calculated, only 15% fit the criteria of "exemplary." Another 11% were considered "deficient." Stone suggested that all public expenditures relating to Board certification be suspended until the NBPTS could prove that its certification process enhanced student learning.

The Stone study elicited responses from the Thomas Fordham Foundation, the Harvard Education Letter, The School Reform News, the Center for Educational Reform, Education Week and of course, the Board itself. The publicly funded Education Commission of the States asked Dominic Brewer, Susan Fuhrman, Robert Linn and Ana Maria Villegas to review Stone's work **(**Furhman, 2002**).** These reviewers were not kind, criticizing the lack of information about the sample, the value-added methodology used in Tennessee, the instrumentation, and several of Stones research procedures. In fact the reviewers were:

> …unanimous in asserting that the conclusions reached by Stone, that "the findings of this study present a serious challenge to NBPTS' claims…" and that "…they suggest that public expenditures on NBPTS certification be suspended…," are completely unsupported by the study. These conclusions severely overreach, considering the methodological limitations identified by reviewers.

When we looked at Stone's data, we quickly saw two issues that needed to be addressed. The first was about consistency of effects for the NBCTs and the second was about reliability of the Tennessee value-added model of assessment.

With regard to the issue of consistency, we wondered why Stone did not make anything out of the fact that in the 23 comparisons of gains in mathematics for NBCTs vs. the average gain made by others in their grade, within their district, 15 (65%) of these comparisons showed the NBCTS to be more effective. In reading, of 29 comparisons, 18 (62%) favored the NBCTs. In language, of 29 comparisons, 16 (55%) favored the NBCTs. In social studies, of 25 comparisons, 14 (56%) favored the NBCTs. And finally, only in science was this trend reversed. Of 17 comparisons among science teachers, only 7 (41%) favored the Board certified teachers. For the most part, in most subject areas, the students of NBCTs scored higher than their peers in the same districts. But we have no way of knowing from this study what metric we are in to determine if the net effect for the NBCTs produced an effect size that was socially significant. Because the NBCTs did not reach the levels that Stone used to judge high levels of accomplishment does not mean that these NBCTs did not produce meaningful gains over their non-NBCT peers.

The second obvious problem with this study was the lack of consistency or reliability of the gains teachers made over different years. This suggests that the value added approach was not effective in reducing the year-to-year variation of teacher test scores due to the near-random assignment of students to teachers. For example, in the year 2000, the students of Teacher #7 in this study scored at 135 percent of the district average in mathematics. But in the year 1999 the students of this NBCT had only scored at 75% of the district average. In one year teacher #7 is a goat, in one year a hero! Teacher #11 showed the same kind of variability on the states' test of language. One year his or her students were 38% above the district average, in another year that teachers' students were at 84% of the districts' average. In the Tennessee value-added system that Stone cites, this same teacher would receive a grade of "A" one year and a grade of "F" another. If a teachers' performance is so dependant on the luck of the draw of students they receive, determining teacher effectiveness by this kind of value added model is seriously flawed.

In fact, Kupermintz (2003; 2002) has seriously questioned the use of value-added models to determine teacher effectiveness. His logical and statistical critique of value-added models suggests

that they are too flawed to be useful in assessment of teacher competency. Gene Glass (2004), a leading statistician and expert on teacher assessment notes:

> Value-added teacher evaluation methods, which attempt to evaluate teachers in terms of the standardized achievement test score gains of their students, are of uncertain validity, have drawn heavy criticism from measurement experts, and raise serious concerns about fairness. They should be opposed in their various forms.

After the release of the Goldhaber and Anthony report reviewed above, Stone (2004) once again argued that NBCTs are not effective. This time he used the argument that statistical significance with such large numbers of students and teachers is irrelevant because the differences between NBCTs and non-NBCTs were actually very small. While that is true, Stone ignores the effect sizes found in the Goldhaber and Anthony study. Those effect sizes appear to average about.10. If we were dealing with a typical norm- referenced standardized test, ESs of .10 would be worth approximately one months gain in a grade-equivalent metric. Relatively small gains in standard scores and scaled scores on standardized tests can indicate relatively large gains in grade equivalents, and grade equivalents are the metric schools and parents often use for judging growth in achievement. This point will be important for interpreting the study that we report below.

Both of Stone's papers call for a suspension of public monies to the NBPTS. We think his data and arguments do not support such a policy change. However his criticism of the expense for determining expertise in teaching through the NBPTS is on more solid ground, an issue that has been raised by others.

Reflecting the continuing criticism of the NBPTS by the Fordham Foundation, a conservative foundation that promotes a market-based approach to defining teacher quality, is a report by Finn and Wilcox (1999). They note that in the 12 years since its inception, the National Board had been unable to prove that NBCTs produced higher achieving students than non-NBCTs. They criticized American businesses for backing the Board-certification process, which they considered to be a losing strategy. They also accused Board-certified teachers and their students as lacking in fundamental knowledge. The solution, in their minds, was to let the market "generate both quality and quantity" (p. 3) of teachers. In this scenario, schools would hire the best teachers they could find, pay them market value and assess their effectiveness by their students' achievement.

One year later, Finn and Wilcox reiterated this stance in an article written for the *Los Angeles Times* (2000). On this occasion, the authors maintained that the Board actually "ignores classroom results" (p. 1). Further criticisms of the Board and its certification process focused on peer review, vague standards, lack of attention to teacher writing skills, questionable items to test subject knowledge, and the ability of candidates to rely on one another for assistance during the year of preparation for certification.

Holland (2002), writing for another conservative foundation, the Lexington Institute, questioned the ability of Board certified teachers to make a difference in student achievement. Holland went on to describe an alternative way in which to attract quality teachers based on a publication of the Fordham foundation (1999) entitled, "The Teachers We Need and How to Get More of Them: A Manifesto." The "manifesto" called for a free market approach in the hiring of teachers. In this approach, teachers would make higher salaries if they could prove that their students made gains in achievement. So too would individuals who taught in subject areas with shortages of teachers. Most importantly, however, the manifesto suggested that future teachers would no longer need to show proof of completed coursework or degree(s) but would simply need to be tested for their "knowledge and skills." Holland perceived the market approach as creating a healthy competition to the Board certification process which he described as being "based largely on theories about self-esteem and child-centered learning" (p.11).

The leaders of North Carolina have supported the National Board's certification program since its inception. In fact, North Carolina's teachers were not only encouraged to seek Board certification, but they were rewarded generously for doing so. The state reimbursed teachers for their Board application fees and upon successful completion of the certification process, provided them a 12% pay raise. Not unexpectedly, the state leads the nation in the numbers of Board certified teachers. As might be expected, therefore, the sizable commitment by the state to the NBPTS has its critics. For example, Leef (2003) wrote for the North Carolina Education Alliance that the state's financial commitment to NBCTs had topped $25 million dollars annually while the Board's standards, he claimed, had little relevance to teaching competency and were loaded with ideas drawn from "progressive education theory." (p. 2). Leef's conclusion, written before the Goldhaber and Anthony study of North Carolina teachers was published, was that no evidence existed to suggest that NBCTs have had a beneficial impact on student learning in the state.

Similarly, M. O. Thirunarayanan (2004) referred to National Board certification for teachers as "a billion dollar hoax" and asked the question, "How much should a nation spend on mediocrity?" Thirunarayanan criticized the Board's standards, referring to them as being representative of entry-level standards for teachers and not adequate for identifying accomplished teachers. In what we think of as a remarkably unrealistic suggestion, Thirunarayanan proposed that Board certification be reserved for those candidates who possess a doctorate in their area of expertise; are able to demonstrate that their students make significantly greater learning gains than did students in other classrooms for at least 5 years; demonstrate they had developed and tested innovations in teaching, learning, and assessment by having been published in a scholarly journal; and have exhibited knowledge of their content area by exceptional performance on "rigorous exams and other assessments."

In our opinion Thirunarayanan's recommendations seem to confuse the university and its academic requirements with K-12 classroom teaching and its norms. His apparent lack of familiarity with teachers and teaching was the first point noted in a response from Jason Margolis (2004), who assumed that Thirunarayanan had never worked with National Board candidates. Margolis directed his attention to the fact that Thirunarayanan had not mentioned the rigorous manner in which candidates are assessed and had assumed that low content standards could be found in all of the Board's content areas. He also noted that Thirunarayanan had not provided any evidence to prove that NBCTs are undeserving of salary raises or opportunities to advance their careers. Margolis also disputed the claim that candidates seek Board certification in an effort to obtain "humongous pay raises" and other incentives. He asserted that NBCTs have proved themselves, through portfolio entries and assessment center activities, to be able to have a positive impact on student learning. Margolis concluded his rebuttal by stating that the National Board certification process, although not perfect, has benefits that are far greater than its drawbacks. He maintained, "Whether teachers certify or not, the process of systematically collecting evidence of impact on student learning in and of itself often furthers teacher development" (p.3).

No Child Left Behind and the quality of teachers. President G. W. Bush signed into law the No Child Left Behind Act (NCLB) on January 8, 2002, the first major education reform to be passed by the federal government in over 40 years. Its intent is to assure that each child in the USA meets the educational standards of the state in which he or she resides. To do that, the law requires that by the 2005-2006 school year all students be taught by "highly qualified teachers."

The NCLB act describes the highly qualified teacher as one who holds a bachelor's degree or higher from a 4-year institution, is fully certified by the state, and can demonstrate competence in his/her subject area. At the elementary level, newly hired teachers will be required to pass a "rigorous state test of subject knowledge and teaching skills in reading/language arts, writing, math

and other areas of the basic elementary curriculum." States are now struggling to develop these rigorous tests of subject knowledge and teaching skill. Some states are considering or have adopted Board certification as evidence that a teacher has satisfactorily demonstrated subject mastery and thus meets the definition of a highly qualified teacher.

The US Secretary of Education, Rod Paige, provides an annual report to Congress entitled "Meeting the Highly Qualified Teachers Challenge." The purpose of the report is to inform the public of the state of teacher quality in America. In the 2003 report, the Secretary acknowledged that research has consistently shown that individual teachers contribute to student achievement. Despite this acknowledgement, however, he questioned how anyone would be able to discern a highly qualified teacher by any means other than by examining the achievement of the teacher's students and is completely silent with regard to the NBPTS and its decades of work (U. S. Department of Education, 2003). This emphasis by the Secretary on student achievement reflects the well-rooted American view about how to judge teacher competence that we discussed above.

We again point out that qualified physicians, lawyers and accountants achieve their designation of competence by virtue of having passed assessments and by their reputation in the field in which they practice. There is little in the way of objective data validating the reputations they earn. Still, we recognize the public's desire to be provided data about teachers, and we understand that those skeptical of the NBPTS need more information about the predictive validity of the Board assessments. This study addresses those concerns directly.

Methods

The purpose of this study was to examine the relationship between National Board certification and student achievement as measured by performance on the Stanford Achievement Test-9th Edition (SAT-9). Fourteen school districts in the state of AZ released SAT-9 scores directly from the Arizona Department of Education for use in this study. Data were examined from the years 1999-2003 for students in grades 2-6. A comparison was made between the adjusted gain scores of students of NBCTs and those of non-NBCTs. The sample of NBCTs consisted of those holding only the Early Childhood or Middle Childhood Generalist certificate(s).

A second part of this study involved the use of on-line questionnaires. The NBCTs and their principals were surveyed to obtain demographic information and opinions about the NBPTS assessment process.

Participants

Teachers. Thirty-five NBCTs residing in the state of Arizona agreed to participate fully in this study. By doing so, they agreed to share their SAT-9 data, if available, complete a questionnaire, and allow their principals to complete one as well. Two other teachers agreed to complete the teacher survey but declined to share SAT-9 data or allow input from their principals.

Principals. A total of 24 principals responded to our survey, representing 24 schools across 14 districts. Some of these principals had supervisory responsibility for more than 1 NBCT. Principals had served at their current locations from 1 to 14 years, serving an average of almost 5 years at their current school. Ninety-one percent of the teachers about whom they were commenting were still employed in their schools.

Districts. Approximately 208,650 elementary students in the state of Arizona attend the 14 school districts that were included in this study. Based on information available from academic year 2001-2002, this is about 36% of Arizona's elementary student population. Eight of the districts (57%) were unified, in that they contained grades K-12. The remaining districts (43%) served only elementary students, typically in grades K-8. The districts ranged in size from 1,249 to 75,359

students, with the average enrollment being more than 19,000 students. The number of elementary students in these districts ranged from 930 to 49,799. At the time the study took place, 62 (77%) of Arizona's EC or MC Generalists were teaching or had taught in one of these 14 districts.

A district was not included in this study if it did not have an EC or MC Generalist in its employ, if all of its EC or MC Generalists did not respond to repeated attempts to be contacted, or if all of the EC or MC Generalists taught at a grade level other than grades 3-6. These grade levels were selected for two reasons. First, the SAT-9 is not mandated in Arizona below grade 2. Although data were available for grade 2 students, no pretest data were available given these students were in grade 1 the year before. The only grade 2 data used were used to calculate pretest scores for those students in classrooms with NBCTs in grade 3. Secondly, students in grades 7 and above change classes throughout the day and therefore would receive only limited instruction from a NBCT.

Procedures

The NBPTS provided us with a list of EC and MC Generalists in the state of Arizona. The combined lists totaled 67 teachers who had received Board certification prior to 2002. In January 2003 the Board forwarded an additional list containing the same information for the individuals who had received certification in November 2002. There were 13 additional individuals in this group, bringing the total of EC and MC Generalists in the state of Arizona to 80.

Each of the NBCTs on the initial list received a mailing that explained the procedures and purpose of the study, as well as a form for updating his/her personal information profile provided by the Board. The teachers were also asked about their interest in participating in the study.

Forty-one (62%) of the teachers on the initial list of 67 responded to the mailing. Of those, 22 (54%) agreed to participate and share their SAT-9 scores. The remainder either had not administered the SAT-9 during the years 1997-2002 or were uninterested in participating in the study. Several wrote comments expressing their views on various topics such as standardized testing and the purpose of the research. (Two of the comments that are relevant for interpreting the findings of this study are included in Appendix A). Unresponsive NBCTs from the first group were re-contacted, while the 13 NBCTs who received their certificates in the fall of 2002 were informed of the study and invited to join. In all, 19 NBCTs (24%) expressed disinterest in participating in the study, 10 (13%) had no test data for the years we were analyzing, and 16 (20%) either could not be located or did not respond to repeated attempts to reach them at home or at work. Thirty-five teachers, about 44% of those who held the kind of Board Certification we were seeking agreed to participate in our study

After the initial mailing, we began contacting representatives of the various school districts to secure permission to survey the teachers and to gain access to the district's SAT-9 data. The districts were promised a small monetary donation in exchange for their cooperation. In all cases permission was granted to conduct this study, however, in one instance, permission was rescinded at a later date. In that situation the director of a suburban Phoenix charter school gave permission for the study but later did not agree to release the school's SAT-9 scores to the researchers.

The Teacher and Principal Surveys.

The teacher survey. An initial version of the teacher survey was piloted with 10 NBCTs who did not hold either the EC or MC Generalist certificate and therefore would not be included in this research. After modifications to the pilot questionnaire, we designed an on-line survey using Remark software (Remark Web Survey, 2002). At the same time, we purchased a domain name and hosting service to accommodate the survey files. The participating NBCTs were given the survey's URL address and directions for how to access the questionnaire. The results were then downloaded

into Excel spreadsheets for analysis. A copy of this questionnaire along with the teachers' responses can be found in Appendix B.

The principal survey. The survey of the principals of the NBCTs was designed to elicit the principals' opinions regarding the Board certified teacher(s) at their school, and their beliefs about assessment in general. The Principals were also asked to describe the manner in which classroom placements were made at their schools. This was of special interest because a number of the NBCTs had reported that needy and difficult students were frequently over-represented in their classrooms and not in the classrooms of their non-NBCT peers (see teacher comment in Appendix C).

As was the case with the teacher survey, the principal survey was designed, piloted, revised and put on-line. All principals were then emailed with the name(s) of their participating NBCT(s) and provided the survey's URL address. A copy of the principal survey and their responses can be found in Appendix C.

Measurement of achievement.

The SAT-9. The SAT-9 is a norm-referenced achievement test published by Harcourt Educational Measurement. The test measures skills in the broad subject areas of reading, math and language using a multiple-choice format. Since 1997, the SAT-9 has been administered yearly to AZ students at varying grade levels. According to The *Administrator's Interpretive Guide* (Stanford Achievement Test-9, 2001), the SAT-9 can "provide national comparative data to assess individual and group performance, and to provide longitudinal data to study changes in performance over time" (p.5). Scaled scores were used in this study because these are the scores that "are suitable for studying change in performance over time" (p.6).

Data Cleaning

The Arizona Department of Education (ADE) created a data disk for us that contained all of the districts' student level SAT-9 data for years 1999-2003 in SPSS format. Seventy separate files (14 districts x 5 grade levels) had to be analyzed and cleaned.

Student data were coded on whether students were in a classroom with an NBCT (or not). For example, in the data year 2001, student data were coded to indicate that the student's teacher for the school year 2000-2001 was an NBCT. These same students were sought out in the data file by student name for the year 2000, the year before they had an NBCT (1999-2000), and were coded as having an NBCT the following year. All other students, those who had not been taught by NBCTs, were coded as the control group in both data sets. Both data files were merged. This process was completed for each of the data years resulting in four distinct data sets 1999-2000, 2000-2001, 2001-2002, and 2002-2003.

We then calculated gain scores for each student. For all of the students coded as "1" under the NBCT variable, the pretest score consisted of the scaled score obtained prior to having an NBCT as a teacher, while the posttest score consisted of the test score obtained during their year with the NBCT. For all other students, those coded as "0," both the pretest and posttest scores were obtained during years in which the student was taught by a non-NBCT. This difference in pre and post-test scores was considered to be the student's one-year gain score.

Students who had gain scores in at least one subject area (reading, mathematics, language, in any year) were kept in the file. All other students, those without gain scores in any subject area, were deleted from the file. This would be the case if students had not taken the test in both data years, or if they had, their scores were not found in any of the 14 districts represented in this study. Because no pretest-posttest analyses could be done, these students' scores were removed from the data set.

With all of the files merged, students coded, and with gain scores calculated in each skill area, it was possible to determine adjusted gains scores (AGSs) of students of NBCTs versus the AGSs of students of non-NBCTs.

Design

Independent Variable. In this study, the independent variable was the National Board certification status of the teachers. The two groups under study consisted of the "treatment group," students of NBCTs, and the control group, the students of non-NBCTs.

Dependent Variable. The dependent variable in this study was the yearly AGS's observed on the SAT-9 by the students of NBCTs and non-NBCTs. The differences in each student's SAT-9 scaled scores from pre to posttest, reported as AGSs, were calculated for each subject area over the period of one year. AGSs of students of NBCTs were compared to AGSs of students of non-NBCTs at each of the four grades, 3-6, for each individual subject area, and for each of the four years of data under analysis.

Qualitative Data. The qualitative data in this study were derived from two single-observation, descriptive surveys, one having been collected from NBCTs and the other from their principals.

Quantitative Data. This study can be classified as an ex-post facto, causal-comparative research design. The scaled scores associated with the SAT-9 were used as a pretest-posttest measure of yearly achievement growth for each student in the sample. The differences in each student's scaled scores from year to year were then reported as gain scores. Gain scores were calculated for each subject area. Following the analyses of covariance of each of the four data sets (1999-2000, 2000-2001, 2001-2002 and 2002-2003), AGS's of students of NBCTs were then compared to AGS's of students of non-NBCTs at each of the four grade levels, 3-6, and for each individual subject area.

The design is to be considered quasi-experimental because we had no control over the placement of students into classrooms and therefore had no control over which students were assigned to classrooms with NBCTs and which were assigned to classrooms with non-NBCTs. Because the students were not placed in classrooms in a random manner, the possibility existed that the groups may have been different prior to the study. For this reason, we used a pretest/posttest design where covariance adjusted gain scores were used to control for the effects that non-random assignment might have had on students' growth over time. The use of the pretest as a covariate reduced the amount of difference or natural variation that could obscure effects within groups, as well as between them. The covariance adjustment makes the two groups we compared more uniform, tending to eliminate bias in the sample. With the kind of data we had to work with analysis of covariance is both a recommended and frequently used procedure (Ferguson and Takane, 1989; Fraenkel and Wallen, 1996; Glass and Hopkins, 1989; Smith and Glass, 1987). But there is no way to guarantee that adjusting the gain scores on the basis of the entering scores of students assigned to these teachers was completely successful in eliminating bias.

Data Analysis

The Surveys. The statistical procedures used to analyze these data varied according to the type of question put forth. Many questions were straightforward Likert items and the percentages or mean numbers of respondents answering at each level of the scale are reported. In some cases more open-ended items were used and the respondents comments were coded and reported. All the original data for teachers and principals are presented in Appendix B and C.

The SAT-9. In order to calculate the AGSs, ANCOVA calculations were performed using a univariate, general linear model in SPSS. The dependent variable, the gain, was determined one subject at a time. The fixed factor was the NBCT status and the covariate was the pretest score for

each of the SAT-9 subject areas. For example, when calculating the ANCOVA for the gains in reading that occurred between 1999-2000, the dependent variable was the reading gain score 2000 while the covariate was the 1999 pretest score in reading. Because the pretest score was used as the covariate, the calculation resulted in AGSs for the NBCT group and the non-NBCT group (see Appendix D for SPSS output).

The effect size (ES) was used to provide a standardized measure of the strength of a relationship and indicated the relative importance of the pretest effect. Effect sizes were computed by dividing the AGSs (in each subject and at each grade level) for each of the two groups by the standard deviation (SD) obtained for that group.

Results

Teacher Survey

Of the 37 NBCTs who agreed to respond to the survey, 34 (92%) actually did so. Twelve of the 34 (35%) had earned the Early Childhood Generalist (EC/Gen) certificate at some time between the years 1995-2002. Another 22 (65%) had been awarded the Middle Childhood Generalist (MC/G) certificate during the same time period. Thirty-three (97%) of the teachers were female and one was male. Of the 24 (71%) who were actively teaching at the time of this survey, 22 (92%) taught in schools located within the metropolitan areas of either Phoenix or Tucson, the state's two largest cities. One participant (4%) taught in a small city located in the rural southeast portion of the state while another (4%) taught in a resort and retirement community located on the state's western border with California. One teacher taught at the preschool level while all others taught in grades K-6. Six of those teaching (25%) taught in multi-age level classrooms. Of the 18 NBCTs (53%) who reported their ethnicity, one teacher (6%) was of Asian/Pacific Island descent while the others reporting were Caucasian (94.4%).

The teacher survey consisted of 27 questions covering three main topic areas: background information about the teachers, views about assessment, and the National Board experience. The full survey and responses to it are included as Appendix B.

Teacher Survey: Background information. The 34 teachers who answered the survey ranged in age from 32 to 64 years. Their median age was 50. They had taught between 10-33 years, with the median number of years experience being 20. Twenty-four of the teachers were still working as classroom teachers, teaching in classes from preschool to grade six. Six of the teachers taught in multi-age classrooms that contained two, three, and in one case, four grade levels! One of the classroom teachers also taught literacy at a major state university. At the time of the survey the remaining teachers were employed as "teachers on assignment outside the classroom," serving, for example, as curriculum specialists, or in new teacher induction, or as resource teachers of various kinds. A few were also in administrative positions.

Almost 80% of the teachers had obtained their teaching degrees from public universities, with about half of these from Arizona's own state universities. Most of the respondents (88%) reported having earned a master's degree. Two held bachelor's degrees, and two had doctorates. Over 80% of the teachers had taken additional coursework after they had obtained their highest degree. Ninety-seven percent of the teachers reported that they participate in professional growth activities on a regular basis. Figure 1 presents the numbers of teachers who participated in specific types of activities during the year immediately before the survey. Most of the "other" responses to the survey involved professional growth activities at the school or district levels.

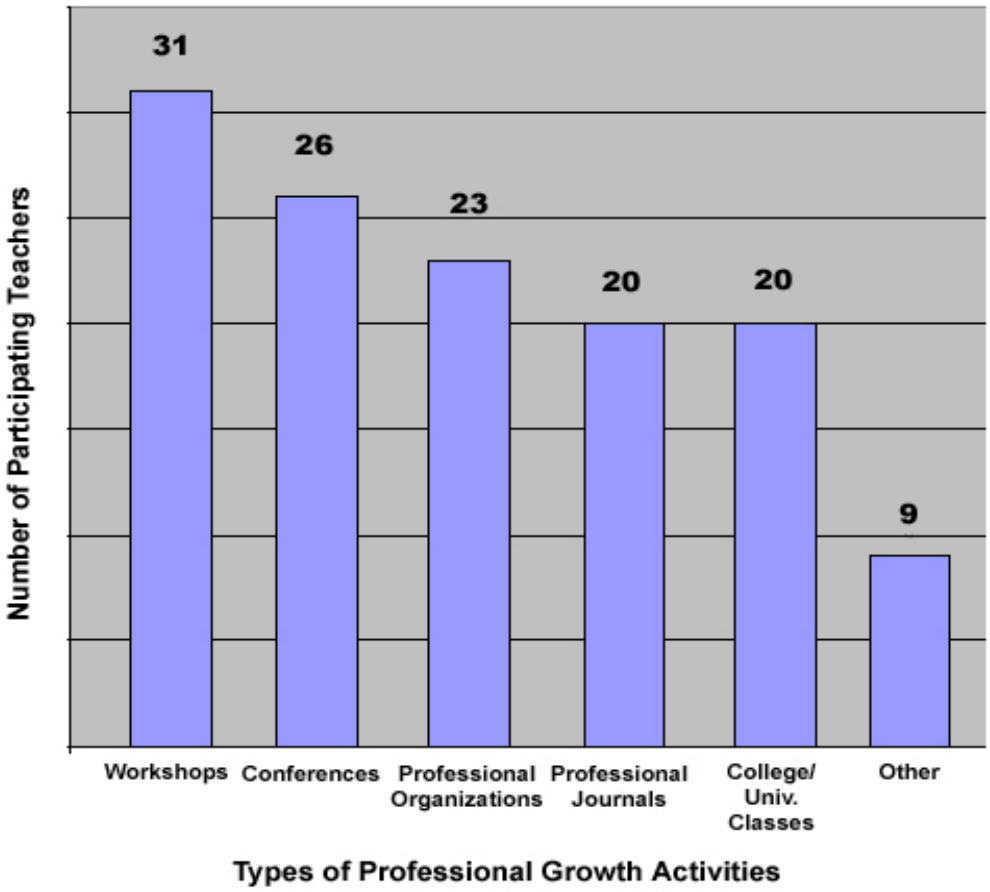

Figure 1. Professional growth activities in the year immediately preceding the survey.

These teachers used a variety of ways to describe their teaching styles. Twenty-four teachers used terms that are unique, such as empowering, multi-sensory, holistic, etc., while others responded with more traditional responses such as hands-on (N=10), student-centered (9), teach to learning styles (8) and structured/ordered (7). Most of the responses indicated that teachers want their students, and in some cases their students' parents, to become actively involved in the learning process.

As a group, these teachers reported having earned a total of 43 endorsements or certificates from the state of Arizona. Over 30% of these endorsements were in the area of special education, including gifted education. Another 19 % had earned endorsements in either English as a Second Language or bilingual education, while another 14% had earned principal certificates. These background characteristics are important to note because the results of this study can easily be explained by the increased professional opportunities in which these teachers participated, and their advanced degree status.

Teacher Survey: Views about assessment. All but one of the NBCTs reported the SAT-9 to be either "very important" or "somewhat important" at their school. When these teachers were asked if they had changed their curricular focus prior to administering the SAT-9, about 55% reported having done so, with most of these including test-taking strategies into their curriculum. Of

the 27 teachers who responded to an item about how SAT-9 results are used, 10 said it was used as a measure of their job performance.

Teachers suggested a variety of other ways in which to measure teacher quality. Direct classroom observation of teaching was the most frequently noted suggestion, however, when combined, the responses of "student achievement" and "student growth over the year" appeared nearly as frequently. When asked how best to measure their students' achievement, over two thirds of the NBCTs used a phrase containing the word "variety" in their responses.

All of the NBCTs who responded to an item about measuring student achievement in their classrooms said that work samples and teacher-made assessments were most frequently used. In addition, about two thirds of the teachers also noted criterion-referenced tests and norm-referenced tests as being common ways of measuring their students' achievement.

Of particular importance for interpreting the data that follow is the fact that ninety percent of the teachers were aware of the criteria or process used to place students in their classrooms. Most classrooms appeared to be a heterogeneous grouping of students based on academic ability, student behavior, gender, special needs, primary language status and ethnicity. In many cases input came from the previous years' teacher(s) and/or the principal. Parents were often allowed to provide input as well. Children were oftentimes assigned to a particular teacher based on a match between their personalities or teaching/learning styles. It is not evident that these teachers received either the "best," or the most difficult to teach students.

Teacher Survey: National Board experience. The NBCTs provided numerous reasons for seeking Board certification. Many found it to be personally, professionally or intellectually challenging. Many others reported it to be a personal or professional growth experience. Some reported that it validated their practice and gave credence to their profession. Many reported that the experience allowed them to bond with other teachers, colleagues, their students and the parents of those students.

Only one teacher reported not having received any type of support when going through the Board certification process. Although most received some type of financial assistance, about two thirds reported also receiving assistance in the form of a university-based program, release time from school or some form of mentoring. Mentoring typically came from other candidates (both past and present), principals, friends, family, colleagues or district teams.

When asked, "In what ways do you think the Board certification process has made you a better teacher?" nearly two thirds of the respondents cited the reflective process as the reason. About one fourth reported that the certification process had resulted in improved student achievement while another 14% reported that they had become more analytical in their approach to teaching.

The NBCTs most common response about the certification process was that they found it to be a significant professional growth experience, both worthwhile and rewarding. Individual teachers reported that it allowed them an opportunity to monitor their own profession and provided them a national platform from which to be heard.

Principal Survey

The Principal Survey consisted of 30 questions covering three broad areas. These included background information about their school and the NBCTs that they supervise, as well as information regarding student and teacher procedures for placement. The second section required the principal to provide archival information and impressions about the NBCTs whom they supervised. The final section included questions regarding the principal's beliefs about such topics as teacher quality, student achievement and the National Board. The full survey and its results are provided in Appendix C.

Principal Survey: Background information. In a question that allowed for multiple responses, most principals reported that students were placed in classrooms by some criterion that made for homogeneity of the classroom, though the form of homogeneity was unspecified. Another common response was that classrooms were structured via a heterogeneous or stratified method. Principals also reported that about one-fourth of the time students were randomly assigned to classrooms, but they also noted that some of the time students were assigned on the basis of parent request, decisions made by a collaborative team of teachers, or on the basis of an hypothesized match of teaching and learning styles. Almost a third of the principals reported that parents were *not* given the opportunity to make any type of teacher request for their child. The remaining principals were divided in regards to the amount of input they allowed parents in choosing a teacher for their child. All but one of the principals reported that classrooms in their school had relatively equal numbers of students.

Principals were also asked the manner in which special needs students were assigned to classrooms. In almost all schools a team approach was used in making these decisions. The teams typically consisted of one or more teachers, the principal, and on some occasions, the parent. Oftentimes the severity of the students' needs was weighted, but in other cases, an attempt was made to place equal numbers of special needs students in each classroom. Contrary to the beliefs of many of the NBCTs, 71% of the principals reported that their NBCTs were assigned the same number of high needs students as their peers. In about 15% of the cases the principals agreed that the NBCT had been assigned more students with special needs.

The typical pattern for a teacher to be assigned to a school was by a team or committee. In about one-third of the cases, teachers were given some degree of choice in the assignment process. After the teachers had been assigned to a school, a decision about which grade level they would teach at was usually made by the principal, but teachers and teams of teachers at a school also had input into those decisions.

Principal Survey: "The National Board Certified Teacher at Your School." Of the principals who allowed parents to request specific teachers, about one-third believed that their NBCT was requested more often than his or her peers. Almost half of the principals had no opinion, hadn't noticed or did not respond to this item. No principals reported the NBCT to be requested less often then his/her peers.

Of the NBCTs who were requested more often than their peers, most were requested due to their prior reputation. These NBCTs were regarded as "outstanding," "good," "experienced," and/or "personable" teachers who provided "quality education" and whose students achieved well.

Thirty-five percent of the principals reported having supervised their NBCT before, during and after the Board certification process. About three-quarters of these principals reported observing changes in the teaching of the NBCTs. They attributed the changes to the Board certification experience. NBCTs were perceived as assuming more of a leadership role, and more willing to try new techniques or take risks. The most frequent response mentioned by the principals, however, involved an increase in the NBCTs reflective practice. Other principals mentioned the teacher's improvements with regard to confidence level, ability to deliver instruction, and skill at differentiating instruction according to students' needs.

Principals were asked to rate their NBCT in comparison to all of the other teachers they had supervised. In only one instance did a principal indicate the NBCT to be the best teacher ever supervised, and also in only one instance was the NBCT judged to be one of the poorest teachers to have been supervised. About 85% of the principals perceived the NBCT to be one of the best teachers ever supervised, though almost ten percent of the principals reported the NBCT to be an average teacher. Figure 2 illustrates these responses.

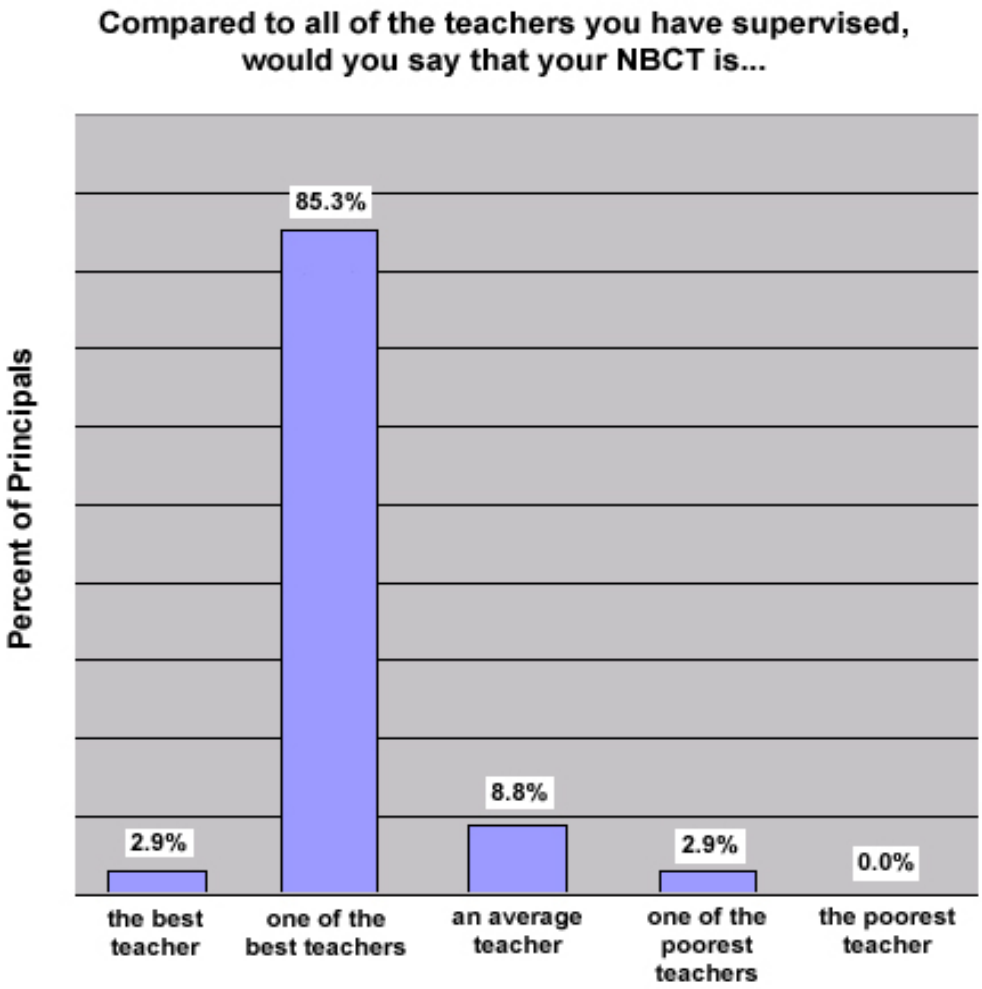

Figure 2. Principals' Ratings of NBCTs Compared to Other Teachers

Principals who rated the NBCT as being the best or one of the best teachers were asked what qualities made these teachers stand out from the rest. About one fourth of the teachers were described as being collaborative, while another 20% of the teachers were described as being organized. Other descriptors suggested that these teachers were dedicated, professional/ethical, motivating/challenging, focused/determined, and communicative. Principals also reported These NBCTs to be leaders. In terms of the classroom instruction provided by these teachers, two-thirds were described as being knowledgeable or having expertise in curriculum and/or, instruction. Another twenty percent or so were noted for their ability to use a variety of teaching methods, while a few others in this group of highly regarded teachers were characterized as meeting the needs of their students, effective in use of time, willing to try new things, and willing to make data-driven decisions.

Principals were also asked to rate their NBCTs in regards to their relationships with colleagues/parents/students, classroom management skills, use of instructional strategies, skills at assessing student learning and planning/goal setting. Six response choices were provided for each item, ranging from 6 for "excellent" to 1 for "unacceptable." The majority of NBCTs were rated as

"excellent" in all areas except "relationships with colleagues." In this case, slightly more NBCTs were rated as "very good" rather than "excellent." Although nearly half of the principals did not respond to an item about changes in the scores received by students of NBCTs, of those who did, only 17% reported having observed any change in the SAT-9 scores of students of their NBCTs.

Other items revealed that the principals perceived their NBCTs as being more involved in professional growth activities than their peers, though 20 percent of the principals saw their NBCTs as no different to their fellow teachers on this dimension. Approximately half of the principals reported their NBCTs to have increased their level of involvement in professional growth activities, but half also believed that their NBCT decreased their level of involvement.

Of the principals who supervised an NBCT during the certification year, about half reported having spent very little or a small amount of time assisting the candidate. A few other principals reported having spent a total of 10-15 hours with their candidate. Only one principal reported having spent as much as two hours per week helping, supervising and cheering the candidate on. Of the 20 principals who responded to this item, almost all believed the time they allocate for this to have been worthwhile.

Principal Survey: "From Your Own Experience." Although one principal noted flaws in the Board certification process, all other principals made positive comments about the Board and/or their NBCT. Over 90% of the principals believe the NBPTS to be contributing to the improvement in teacher quality. The other's had no opinion or didn't know, but did not disagree with the perception of the majority. Furthermore, when asked if they believe that the NBPTS is contributing to improvements in student achievement, 70% of the principals believed this to be true. The other 30% had no opinion or didn't know, but did not disagree with the perception of the majority.

In this section of the questionnaire an effort was made to determine the principals' beliefs about the importance of certain variables that might influence a student's SAT-9 scores. The results were not unusual—teacher quality, parent involvement, student ability, student socio-economic level, and so forth. Principals were divided, however, in their beliefs about the best ways to measure teacher quality. More principals reported direct observation of teaching to be important for measuring teacher quality than tests of student growth and achievement, though the two were almost mentioned an equal number of times. Just over half of the principals believed that student achievement is best measured using a variety of measures, while about one-forth of the principals reported that achievement should be measured over time rather than as a static snapshot of student achievement at a single point in time. The majority of these principals were not hostile to using standardized norm-referenced tests to measure teacher quality.

In sum, these principals added little to our knowledge of how students are placed with teachers, but responded with nothing to concern us that students were placed with NBCTs in a way that might be biased in their favor. Furthermore, the principals informed us that Board certified teachers seemed to be higher quality instructors than their average teachers, mirroring some of the survey research reviewed previously. However, if these principals' ratings are to be trusted, a few false positives among the NBCTs may have been identified, an issue we brought up earlier. On the other hand, there is no reason to believe that the validity of these principals' ratings is high enough for us to trust that they can accurately identify such individuals.

SAT-9 Analyses

The procedures and calculations described in the methods section resulted in yearly reports of the "Adjusted Gain Score Statistics." These can be found in Tables 1-4. These tables were created to reflect the results of the ANCOVA by data set (1999-2000, 2000-2001, 2001-2002 and 2002-2003), grade level (3, 4, 5, and 6) and subject area (reading, math and language). For each year, grade,

and subject, we report the numbers (N) of students in each of the two groups being studied, those of NBCTs and those of non-NBCTs. The AGSs and the standard deviations are reported for each group as well. Differences between AGSs are reported as significant at the $p < .05$ level. Finally, each table indicates the effect sizes (ES) for each group, and the difference between each group's ES.

Results for 1999-2000 (See Table 1). Between the years 1999-2000, students of NBCTs made greater gains than students of non-NBCTs in 75.0% (9/12) of the total comparisons. Gains were significant at the $p < .05$ level in 33.3% (3/9) of these comparisons. There were two instances in which students of non-NBCTs outperformed those of NBCTs, but the differences were not statistically significant. In one case, the AGSs were equal across groups.

In 83.3% (10/12) of the total cases, ESs were larger for students of NBCTs than for students of non-NBCTs. The ESs in favor of NBCTs averaged .134 in reading, .352 in mathematics and .125 in language. Glass (2002) helps to put the magnitude of these ESs into a perspective that differs from Cohen's (1988) often cited work on the interpretation of ESs. Glass showed that an ES of 1.0 is approximately equivalent to one academic year's growth on a typical standardized test. Since an academic year is ordinarily ten months in length, an ES of +.10 is roughly equal to a one-month advantage on the grade equivalent scale of a standardized test. In comparison to the students of non-NBCTs, in the academic year 1999-2000, students of NBCTs gained about one and a third months' more in reading achievement, three and half months more in mathematics acheivement, and one and a quarter months' more in language compared to the students of non-NBCTs.

Results for 2000-2001 (See Table 2). Between the years 2000-2001, students of NBCTs made greater gains than students of non-NBCTs in 75.0% (9/12) of the total comparisons. Gains were significant at the $p < .05$ level in 33.3% (3/9) of these comparisons. There were three instances in which students of non-NBCTs outperformed those of NBCTs, but the differences were not statistically significant.

In 75.0% (9/12) of the total cases, ESs were larger for students of NBCTs than for students of non-NBCTs. The ES in favor of the students of the NBCTs averaged .149 in reading, about a one and one half months advantage; .048 in mathematics, about a half months advantage; and .21 in language, representing over a two months advantage.

Results for 2001-2002 (See Table 3). Between the years 2001-2002, students of NBCTs made greater gains than students of non-NBCTs in 58.3% (7/12) of the total comparisons. None of these gains was significant at the $p < .05$ level. There were five instances in which students of non-NBCTs outperformed students of NBCTs, but none of these gains was significant at the $p < .05$ level either.

In 66.6% (8/12) of the total cases, ESs were larger for students of NBCTs than for students of non-NBCTs. The ES in favor of the students of the NBCTs averaged .04 in reading, under a half-months advantage; they averaged .109 in mathematics, about a one months advantage; and they averaged -.038 in language, representing between a quarter and a half-months advantage in favor of the students of non-NBCTs.

Table 1
Adjusted Gain Score Statistics (1999-2000)

Grade 3

	Reading					Math					Language				
	N	Mean Gain*	S.D.	Sig.**	**ES**	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**
Non-NBC	14506	32.3	26.7		**1.210**	15231	28.5	28.6		**0.997**	15207	29.2	28.2		**1.035**
NBCT	113	36.0	26.7	No	**1.348**	121	34.5	24.5	Yes	**1.408**	123	30.3	25.2	No	**1.202**
Difference		3.7			**0.138**		6			**0.708**		1.1			**0.167**

Grade 4

	Reading					Math					Language				
	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**
Non-NBC	15487	28.4	24.1		**1.178**	15962	32.9	26.2		**1.256**	15955	21.6	25.9		**0.834**
NBCT	184	28.4	22.2	No	**1.279**	186	30.1	24.5	No	**1.229**	190	21.5	23.2	No	**0.927**
Difference		0			**0.101**		-2.8			**-0.027**		-0.1			**0.093**

Grade 5

	Reading					Math					Language				
	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**
Non-NBC	15550	12.4	22.7		**0.546**	16021	24.4	24.6		**0.991**	15987	11.3	23.0		**0.491**
NBCT	77	19.4	23.2	Yes	**0.836**	82	36.4	21.1	Yes	**1.725**	81	15.9	22.4	No	**0.710**
Difference		7			**0.290**		12			**0.734**		4.6			**0.219**

Grade 6

	Reading					Math					Language				
	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**
Non-NBC	11752	11.0	19.9		**0.553**	11998	21.0	22.6		**0.929**	11930	13.0	21.9		**0.594**
NBCT	79	11.6	20.8	No	**0.558**	79	23.5	25.5	No	**0.922**	79	16.6	27.1	No	**0.613**
Difference		0.6			**0.005**		2.5			**-0.007**		3.6			**0.019**

*Adjusted gain scores (AGSs) calculated controlling for pretest scores

** Difference between adjusted gain scores is significant at a $p < .05$ level

Table 2
Adjusted Gain Score Statistics (2000-2001)

Grade 3

	Reading					Math					Language				
	N	Mean Gain*	S.D.	Sig.**	**ES**	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**
Non-NBCT	15261	31.2	26.5		**1.177**	16116	25.0	28.8		**0.868**	16093	28.9	28.2		**1.025**
NBCT	169	37.4	28.2	Yes	**1.326**	173	25.1	26.0	No	**0.965**	172	31.7	29.1	No	**1.089**
Difference		6.2			**0.149**		0.1			**0.097**		2.8			**0.064**

Grade 4

	Reading					Math					Language				
	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**
Non-NBCT	16313	29.2	24.5		**1.192**	16786	30.4	25.8		**1.178**	16742	20.1	25.4		**0.791**
NBCT	154	28.3	24.5	No	**1.155**	154	28.1	24.3	No	**1.156**	154	20.3	22.8	No	**0.890**
Difference		-0.9			**-0.037**		-2.3			**-0.022**		0.2			**0.099**

Grade 5

	Reading					Math					Language				
	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**
Non-NBCT	16682	12.7	22.4		**0.567**	17294	24.0	24.1		**0.996**	17159	11.3	22.4		**0.504**
NBCT	89	15.0	19.9	No	**0.754**	84	27.6	24.0	No	**1.150**	87	15.2	21.8	No	**0.697**
Difference		2.3			**0.187**		3.6			**0.154**		3.9			**0.193**

Grade 6

	Reading					Math					Language				
	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**
Non-NBCT	12642	12.3	20.0		**0.615**	12931	20.7	22.6		**0.916**	12773	13.0	22.0		**0.591**
NBCT	64	17.0	18.7	Yes	**0.910**	63	20.4	23.2	No	**0.879**	65	20.1	18.7	Yes	**1.075**
Difference		4.7			**0.295**		-0.3			**-0.037**		7.1			**0.484**

*Adjusted gain scores (AGSs) calculated controlling for pretest scores

** Difference between adjusted gain scores is significant at a $p < .05$ level

Table 3
Adjusted Gain Score Statistics (2001-2002)

Grade 3

	Reading					Math					Language				
	N	Mean Gain*	S.D.	Sig.**	**ES**	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**
Non-NBCT	15339	32.2	26.5		**1.215**	15973	26.4	28.9		**0.913**	15951	30.0	28.3		**1.060**
NBCT	144	32.7	25.8	No	**1.267**	150	26.7	28.8	No	**0.927**	153	27.1	30.6	No	**0.886**
Difference		0.5			**0.052**		0.3			**0.014**		-2.9			**-0.174**

Grade 4

	Reading					Math					Language				
	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**
Non-NBCT	16178	28.2	23.9		**1.180**	16640	30.9	25.9		**1.193**	16552	20.0	25.4		**0.787**
NBCT	114	31.1	23.8	No	**1.307**	116	30.8	26.7	No	**1.154**	113	17.8	21.2	No	**0.840**
Difference		2.9			**0.127**		-0.1			**-0.039**		-2.2			**0.053**

Grade 5

	Reading					Math					Language				
	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**
Non-NBCT	17029	12.2	22.3		**0.547**	17506	25.3	24.1		**1.050**	17384	11.6	22.6		**0.513**
NBCT	60	9.4	22.5	No	**0.418**	61	28.0	23.9	No	**1.172**	62	12.9	24.4	No	**0.529**
Difference		-2.8			**-0.129**		2.7			**0.122**		1.3			**0.016**

Grade 6

	Reading					Math					Language				
	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**
Non-NBCT	14421	12.3	20.0		**0.615**	14821	20.7	22.9		**0.904**	14651	13.8	21.9		**0.630**
NBCT	81	11.2	15.5	No	**0.723**	80	25.1	20.2	No	**1.243**	78	15.1	25.8	No	**0.585**
Difference		-1.1			**0.108**		4.4			**0.339**		1.3			**-0.045**

*Adjusted gain scores calculated controlling for pretest scores

** Difference between adjusted gain scores is significant at a $p < .05$ level

Results for 2002-2003 (See Table 4). Between the years 2002-2003, students of NBCTs made greater gains than students of non-NBCTs in 83.3% (10/12) of the total comparisons. Gains were significant at the $p < .05$ level in 50.0% (5/10) of these comparisons. There were two instances in which students of non-NBCTs outperformed those of NBCTs, but the differences were not statistically significant.

In 75.0% (9/12) of the total cases, ESs were larger for students of NBCTs than for students of non-NBCTs. The ES in favor of the students of the NBCTs averaged .225 in reading, about a two and a quarter months advantage; .065 in mathematics, over a half month's advantage; and .047 in language, representing almost a half months advantage.

Table 4
Adjusted Gain Score Statistics (2002-2003)

Grade 3

	Reading					Math					Language				
	N	Mean Gain*	S.D.	Sig.**	**ES**	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig	**ES**
Non-NBCT	15541	31.2	26.2	Yes	**1.191**	16275	27.3	29.4	Yes	**0.929**	16144	30.0	28.4	No	**1.056**
NBCT	105	38.6	25.3		**1.526**	118	33.7	27.9		**1.208**	117	34.6	28.9		**1.197**
Difference		7.4			**0.335**		6.4			**0.279**		4.6			**0.141**

Grade 4

	Reading					Math					Language				
	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**
Non-NBCT	16103	28.0	24.1	Yes	**1.162**	16582	31.4	26.0	Yes	**1.208**	16529	19.6	25.8	Yes	**0.760**
NBCT	123	32.2	22.4		**1.438**	127	39.1	26.8		**1.459**	128	25.2	30.4		**0.829**
Difference		4.2			**0.276**		7.7			**0.251**		5.6			**0.069**

Grade 5

	Reading					Math					Language				
	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**
Non-NBCT	16565	12.8	22.5	No	**0.569**	17020	25.3	24.3	No	**1.041**	16905	11.3	22.5	No	**0.502**
NBCT	40	14.4	25.0		**0.576**	41	24.1	29.7		**0.811**	43	6.8	20.9		**0.325**
Difference		1.6			**0.007**		-1.2			**-0.230**		-4.5			**-0.177**

Grade 6

	Reading					Math					Language				
	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**	N	Mean Gain	S.D.	Sig.	**ES**
Non-NBCT	12743	12.1	19.8	No	**0.611**	12984	19.9	22.4	No	**0.888**	12914	13.3	22.1	No	**0.602**
NBCT	84	14.0	15.7		**0.892**	84	23.0	26.9		**0.855**	84	16.6	22.0		**0.755**
Difference		1.9			**0.281**		3.1			**-0.033**		3.3			**0.153**

*Adjusted gains scores calculated controlling for pretest scores

** Difference between adjusted gain scores is significant at a $p < .05$ level

Results across the years from 1999-2003. In all, students of NBCTs outperformed students of non-NBCTs on 72.9% (35/48) of the measures during years 1999-2003. Students of non-NBCTs outperformed the students of NBCTs in 25% (12/48) of the comparisons. In one case the mean gains were equal for both groups.

In 11 of the 35 comparisons (31.4%) in which students of NBCTs outperformed students of non-NBCTs, the differences were statistically significant at the $p < .05$ level. In the cases in which students of non-NBCTs outscored the students of NBCTs, none of the differences were significant. In other words, of the statistically significant findings, students in classrooms with NBCTs outperformed students in classrooms with non-NBCT teachers 100% of the time.

In 75% (36/48) of the total cases, ESs were larger for students of NBCTs than for students of non-NBCTs. For the four years studied the effect sizes on the SAT-9, averaged across curriculum areas, were .203, .135, .037 and .112, yielding an overall average ES of about .122, indicating over one months gain per year on this standardized achievement test.

Results across the three subject areas from 1999-2003. In reading, students of NBCTs outperformed the students of non-NBCTs in 12 of the 16 comparisons (75.0%). Five of these 12 comparisons (41.7%) were significant at the $p < .05$ level. In three comparisons the students of non-NBCTs outperformed the students of NBCTs, but in none of these cases were the differences statistically significant. In one instance, there was no difference in adjusted gains between the two groups.

In math, students of NBCTs outperformed the students of non-NBCTs in 11 of the 16 comparisons (68.8%). Four of these 11 comparisons (36.4%) were significant at the $p < .05$ level. In five comparisons the students of non-NBCTs posted greater adjusted gains than did the students of NBCTs, but in none of these cases were the differences statistically significant.

In language, students of NBCTs outperformed the students of non-NBCTs in 12 of the 16 comparisons (75.0%). Two of these 12 comparisons (16.7%) were significant at the $p < .05$ level. In four comparisons the students of non-NBCTs posted greater adjusted gains than did the students of NBCTs, but in none of these cases were the differences statistically significant.

For all of the data years combined, adjusted gain scores of students of NBCTs exceeded those of students of non-NBCTs by an average of 2.39 scaled score points in reading, 3.11 scaled score points in math and 1.86 scaled score points in language. When the differences in adjusted gain sores of each of these content areas were averaged, the resulting mean adjusted scaled score gain for all subjects and across all data years was 2.45 points. Students of NBCTs averaged 2.45 points greater adjusted gains in scaled scores on the SAT-9 per year than did students of non-NBCTs.

Analysis of Effect Sizes. Of considerable importance are the differences noted in effect sizes between the two groups. The mean difference in effect sizes across all of the subject areas and for all of the years for which we had data was just over +0.12. An effect size of this magnitude indicates that the effect of having a National Board Certified Teacher for students is not trivial. The students of NBCTs have over a one month advantage in achievement in comparison to the students taught by non-NBCTs.

Of interest is that the difference in effect sizes between the two groups was greater in the areas of math and reading than in language. Personnel at the Arizona Department of Education believe that Arizona's academic standards are more closely aligned to the skills measured on the SAT-9 in reading and math than to the skills measured in the area of language (Laczko-Kerr and Berliner, 2002). The rationale for this assumption is that the SAT-9 does not require students to provide a writing sample in any of the language subtests. The SAT-9 tests competency in language

via a multiple-choice response format, stressing such conventions as punctuation/capitalization, the ability to manipulate words and phrases or to recognize correct sentence structure, etc. (see: http://www.ade.state.az.us/standards/stanford9/stanford9factsheet.asp). On the other hand, the Arizona language standards emphasize the need for students to produce writing samples at an appropriate developmental level from kindergarten to grade 12 (see: http://www.ade.state.az.us/standards/language-arts/std2.pdf). Therefore, if NBCTs teach to the standards and not to the test, it might be expected that their adjusted gains in language would be smaller than in reading or math. (As we noted in the discussion of the surveys, many NBCTs reported having undergone the certification process in an effort to better align their teaching to state standards. Others noted that they changed their curricular focus prior to administering the SAT-9 in an effort to put greater emphasis on the standards. Clearly, the Board emphasizes standards as being the foundation upon which good teaching is based).

Following Glass's analysis and assuming that an academic year is ten months in length, then this study has shown that on average, students of NBCTs made over 1.3 months greater gains per year in reading than did students of non-NBCTs. In math, the students of NBCTs averaged over 1.4 months greater gain than their peers. In the subject/content area of language, the gain attributed to having studied with a NBCT averaged over three fourths of one month's growth. When ESs are averaged across years and across subjects, we get an overall rule of thumb about what might be expected should these results be generalizable: Students of NBCTs averaged over 1.2 months greater gain than did students in classrooms taught by non-NBCTs. Even though the students of NBCTs score only slightly higher on the scaled score metric of the SAT-9, such gains convert into a non-trivial monthly gain in academic school subjects.

Discussion and Conclusions

The charge has sometimes been made that NBCTs have easier-to-teach students than other teachers and that factor could account for any gains they might demonstrate. But Goldhaber and Anthony found no evidence to believe that the slight advantages that the NBCTs had with regard to the characteristics of the student body they taught made any statistical difference in the results.

In our study three issues insure that we also can disregard this factor. First was the design decision to use peer (within district) teachers as the comparison group, second was the use of covariance analysis to control for non-random factors in entering ability, and third was the collection of information about these issues from the principal survey. All our sources of data suggest that the NBCTs we studied were not teaching easier-to-teach students, but if they were, it was a factor for which we had designed statistical controls. In fact, the teacher survey we used suggested that NBCTs might actually have harder-to-teach children than other teachers in their own schools (see comment, Appendix A). Our own guess is that a systematic positive or negative bias with regard to the students NBCTs teach is likely to vary from one school and one district to another. Without random assignment we can never be sure. But given the assurances of the principals and from our discussions with the teachers, we are confident that no systematic bias in favor of the NBCTs influenced our results.

It is also important to note that many NBCTs do not believe that the SAT-9, nor any other standardized test, will ever adequately assess what they do in their classes. In one study (Rapp, 2001) an overwhelming majority of the NBCTs believed that the state's standardized tests are harming education. Many of the NBCTs we talked to purposefully do not teach in order that their students receive good scores on the test (though many do). Some of these teachers, therefore, did not participate in this study because they found our assessment of them by means of a standardized test

to be demeaning. (See Appendix A for one such comment). Our sample of NBCTs, therefore, consisted of those who were willing to engage in this research, but it was made clear to us that the majority of the sample we used were not enthusiastic supporters of the research design.

What we did learn from this sample of NBCTs was quite similar to what was learned by Goldhaber and Anthony (2004). Board certified teachers have effects on student achievement beyond that produced by non-Board certified teachers. Like Goldhaber and Anthony the effects on the achievement test used as an outcome measure for the research appear to be small, but once again, as in their study, the effect size may be quite compelling. On average, over the academic years studied and in comparison to non NBCTs, the Board certified teachers in our study produced over three- fourths of one months gain (Language) to one and one-third months gain (Reading) on the standardized achievement tests that Arizona uses to measure student progress. Achievement gains for the students of NBCTs across the various subject matter areas averaged over 1.2 months more than for the students of peer teachers who were not certified. It is as if the NBCTs were able to get in about 25 more days of instruction in the typical 180 day calendar, or teach for about 205 days instead of 180 days each year.

Given the weakness in the studies that showed no relationship between Board certification and student achievement (Stone, 2002 and Stephens, 2003) and the strengths of the Bond, Smith, Baker, & Hattie (2000) study (showing deeper student classroom work) as well as the Goldhaber and Anthony (2004) study and our own, the preponderance of the evidence suggest that the students of NBCTs achieve more.

Perhaps some of the critics of the NBPTS, such as Podgursky, might reconsider his (2001) remark that :

> …no rigorous study has ever been undertaken to determine whether the students of board-certified teachers actually learn more than students of an average teacher in the workforce (or teachers who have failed the board assessment), where student achievement is measured by a state assessment or a standardized objective exam.

Perhaps Finn will also now be placated, and enthusiastically support the idea that NBCTs are deserving of extra pay. Given the weight of the extant research it would seem to meet criteria that he set for such support. He and Wilcox (2000) noted:

> If, in fact, the board could guarantee that teachers who earn its credential do an outstanding job of imparting skills and knowledge to their pupils, generous rewards to those teachers would make sense. Regrettably, the board cannot make that claim. In fact, the board ignores classroom results.

Mentioned earlier was that there could be a higher-than-average number of false negative and false positives among those taking the certification test battery. We believe that without extensive and very expensive classroom observations of teaching this will always be the case. Too much teacher knowledge is tacit, knowledge-in-action, and thus extremely hard to assess with paper and pencil instruments or a rubric that assess what a teacher says about his or her own video-taped performance. Too much of teaching depends on context (Moss et al., 2004), making the generalizability of judgments from assessments like those designed by the NBPTS quite difficult. Nevertheless, the NBPTS certification process seems a reasonable compromise between a) prohibitively expensive classroom observations and analyses of teaching, requiring data collection over many different days of teaching, with different observers on different days, and b) a very cheap and quick paper and pencil test of teacher competence, with the likelihood of seriously limited validity. Now that we have the studies of Bond et al., Goldhaber and Anthony, and our own, we can

see that the NBPTS seems to have gotten it right. Their time-consuming performance tests are of considerable (but not out-of-reach) costs, and have proven to have four kinds of validity.

The Board assessments have shown construct validity, as demonstrated by Bond, Smith, et al. who found that teachers identified as expert teachers through the Board assessments were, in fact, conducting their classrooms as experts were hypothesized to do. The assessment battery is aligned with the construct of expertise in teaching.

The predictive or criterion validity of the test was demonstrated in the Goldhaber and Anthony study, and our own. The Board assessments identify teachers whose students produce more learning per year than do the students of non-board certified teachers, including those that tried but did not pass the exams. Because the design of the studies by Bond, Smith et al., and by Goldhaber and Anthony looked at those who took the Board examinations and failed, and found significant differences between them and those who received Board certification, the assessments have proven capable of making some very fine grained discriminations among teachers.

The content validity of the assessments is certainly acceptable. This conclusion is reached from the reports of the committees that make up the tests in each of the 27 areas for which certification is offered, and from the reports of candidates who have taken the tests.

Finally, with so many Board candidates noting that they changed their teaching as a function of preparing for the test, we see evidence of catalytic validity. The test is apparently driving changes in teaching even among those that take the tests and fail in obtaining Board certification.

One major issue unaddressed in this paper is the costs of the exam to a state and its taxpayers. The decision to offer money to teachers for attempting the Boards, and to then provide annual supplements for successful NBCTs, is a problematic one. State revenues and educational values influence these decisions (for example, see Kearney's (2000) analysis promoting fiscal support for the State of New York). We do not comment on these issues. Most critics of the NBPTS have brought up the issue of cost when pointing out that there was no evidence that NBCTs made a difference in the achievement of school children. The critics asked whether the financial commitments to NBCTs were worth the investment without evidence that they do influence student achievement in a positive way. But persuasive evidence now exists that NBCTs do influence their students' achievements in positive ways. For those in the world of business who talk about pay for performance, and the leadership of the National Alliance for Business who support investing in the professional growth of teachers through Board certification, the issue now seems much clarified. It seems to boil down to this: How big a stipend should states pay their NBCTs who we now know demonstrate excellence in the performance of their duties (as noted in the Bond, Smith, et al., study), and whose teaching yields significantly higher productivity (as evidenced by greater growth in student achievement).

In this study we provided evidence that elementary level NBCTs in the state of Arizona are judged to be superior teachers and leaders in their field by their supervisors, and do, on average, raise student achievement more over the course of a year than do non-NBCTs. The amount they raise student achievement, compared to their peer teachers, is socially as well as statistically significant, amounting on average to over one month's more growth for students. The weight of the current evidence suggests that the NBPTS conducts a certification program that works as intended and that the state of Arizona might want to consider supporting its NBCTs instead of ignoring them.

Note

This study is adapted from the dissertation of the first author, *National Board Certified Teachers and Student Achievement*, completed at the College of Education, Arizona State University, May, 2004. Partial support for this work was provided by a research grant from the National Board for Professional Teaching Standards. We are grateful to them for the total independence they granted us to do this research. The views expressed in this paper, therefore, are solely those of the authors.

References

Ballou, D. (1998, June 10). Some unanswered questions concerning National Board Certification of teachers. *Education Week on the WEB,* Retrieved June 25, 2004 from http://www.edweek.org/ew/ew_printstory.cfm?slug=39ballou.h17

Beldon, N. (2002). California teachers' perceptions of National Board Certification. Santa Cruz, CA: Center for the Future of Teaching and Learning.

Berg, J. H. (2003). Improving the quality of teaching through National Board Certification: Theory and Practice. Norwood, MA: Christopher-Gordon publishers

Berliner, D. C. (1994a). Expertise: The wonders of exemplary performance. In John N. Mangieri and Cathy Collins Block (Eds.), *Creating powerful thinking in teachers and students.* Ft. Worth, TX: Holt, Rinehart and Winston.

Berliner, David C. (1994b). Teacher expertise. In Torsten Husèn and T. Neville Postlethwaite (Eds.), *The international encyclopedia of education* (2nd ed.)(Vol. 10, pp. 6020-6026). London: Pergamon.

Bond, L. (2001). *Defrocking the National Board. A reply to Podgursky.* Greensboro, North Carolina: Department of Educational Research Methodology & Center of Educational Research and Evaluation, University of North Carolina.

Bond, L., Jaeger, R. M., Smith, T. & Hattie, J. A. (2000). *Accomplished teaching validation study*. Retrieved March 3, 2001, from http://new.nbpts.org/Press/exec_summary.pdf

Bond, L., Smith, T., Baker, W. K. & Hattie, J. A. (2000, September). *The certification system of the National Board for Professional Teaching Standards: A construct validity study*. Greensboro, North Carolina: Center for Educational Research and Evaluation, The University of North Carolina at Greensboro.

Buday, M., & Kelly, J. (1996). National Board certification and the teaching professions commitment to quality assurance. *Phi Delta Kappan, 78* (3), 215–219.

Burroughs, R., Schwartz, T. A. & Hendricks-Lee, M. (2000, February). Communities of practice and discourse communities: Negotiating boundaries in NBPTS certification. *Teachers College Record*, *102* (2), 344-374.

Carnegie Forum on Education and the Economy. (1986, May). *A nation prepared: Teachers for the 21st century, the report of the Task Force on Teaching as a Profession*. Washington, DC: Author.

Castor, B. (2001). *National Board certification: Vital, rigorous and market driven*. Retrieved July 2, 2004, from http://www.nbpts.org/pdf/castor_commentary.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates: NJ.

Dagenhart, D. B. (2002). *Comparing the wants and needs of National Board Certified with non-National Board Certified middle school teachers for personal job success and satisfaction*. Unpublished doctoral dissertation, College of Education, University of North Carolina, Chapel Hill.

Ferguson, G. A. (1998). Teachers' perceptions and expectations and the black-white test score gap (273-317). In C. Jenks and M. Phillips (Eds.), *The black-white test score gap*. Washington, DC: Brookings Institution Press.

Ferguson, G.A. & Takane, Y. (1989). *Statistical analysis in psychology and education*. New York, NY: McGraw Hill.

Finn, Jr., C. E. & Wilcox, D. D (1999, August 9). *Board games: Failure of the National Board for Professional Teaching Standards to accomplish objective of improving quality of teaching in the US; Business backs a losing education strategy*. Retrieved June 30, 2004, from http://www.edexcellence.net/foundation/publication/publication.cfm?id=161

Finn, Jr., C. E. & Wilcox, D. D. (2000, January 13). *Teachers should be graded on how well their students are learning*. Retrieved June 21, 2004, from http://www.edexcellence.net/foundation/publication/publication.cfm?id=102

Fordham Foundation (1999). The teachers we need and how to get them. Retrieved June 25, 2004 from http://www.edexcellence.net/foundation/publication/publication.cfm?id=16

Fraenkel, J.R. & Wallen, N.E. (1996). *How to design and evaluate research in education* (3rd Ed.). New York, NY: McGraw Hill.

Fuhrman, S. (2002, September 13). *Synthesis of reviews of "The value-added achievement gains of NBPTS-certified teachers in Tennessee: A brief report.*" Retrieved on July 10, 2003, from http://ecs.org/html/special/nbpts/Panel Report.htm

Glass, G. V (2002). Teacher characteristics. In A. Molnar (Ed.), *School reform proposals: the research evidence* (Chapter 8). Retrieved June 30, 2004 from http://www.asu.edu/educ/epsl/EPRU/documents/EPRU 2002-101/Chapter 08-Glass-Final.htm

Glass, G. V (2004, April). *Teacher Evaluation* (Policy Brief, EPSL-0401-112-EPRU). Tempe, AZ: Education Policy Studies Laboratory, College of Education, Arizona State University. Retrieved June 17, 2004, from http://edpolicylab.org

Glass, G.V & Hopkins, K.D. (1989). *Statistical methods in education and psychology* (3rd Ed.). Needham Heights, MA: Allyn and Bacon.

Goldhaber, D. (2002). The mastery of good teaching: Surveying the evidence on student achievement and teacher characteristics. *Education Next, 2* (1), 50-55.

Goldhaber, D. & Anthony, E. (2004). Can teacher quality be effectively assessed? Retrieved on June 24, 2004, from www.crpe.org/workingpapers/pdf/NBPTSquality_report.pdf

Goldhaber, D. D. & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, *22* (2), 129-145.

Goldhaber, D. D. & Brewer, D. J. (1996). Evaluating the effects of teacher degree level on educational performance. *Developments in School Finance* 1996. Retrieved July 18, 2001, from http://nces.ed.gov/pubs97/975351.html

Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy, 100*(1), 84-117.

Haycock, K. (1998, Summer). Good teaching matters… a lot. Retrieved July 2, 2004, from http://www2.edtrust.org/NR/rdonlyres/0279CB4F-B729-4260-AB6E-359FD3C374A7/0/k16_summer98.pdf

Holland, R. (2002, December). *National teacher certification: Advancing quality or perpetuating mediocrity?* Retrieved June 30, 2004 from http://www.lexingtoninstitute.org

Indiana National Standards Board (2002, Spring). *Status of National Board Certified Teachers in Indiana*. Retrieved June 30, 2004 from http://www.nbpts.org/pdf/indianapaper.pdf

Iovacchini, L. C. (1998). *National Board for Professional Teaching Standards: What teachers are learning.* Unpublished doctoral dissertation, College of Education, University of South Carolina, Columbia.

Kearney, C. P. (2000, November 8). National board certification: An underutilized resource for New York State? Paper presented at the Education Finance Research Consortium, Symposium on the Teaching Workforce, Albany, New York. Retrieved June 30, 2004 from: http://www.albany.edu/edfin/symp2000/KearneyW3.pdf

Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee Value Added Assessment System. *Educational Evaluation and Policy Analysis, 25*(3), 287-298.

Kupermintz, H. (2002). Value-Added Assessment of Teachers, In Molnar, A. (Ed.) *School Reform Proposals: The Research Evidence.* Greenwich, CT: Information Age Publishing, Inc.

Laczko-Kerr, I., & Berliner, D. C. (2002, September 6). The effectiveness of "Teach for America" and other under-certified teachers on student academic achievement: A case of harmful public policy, *Education Policy Analysis Archives*, *10* (37). Retrieved June 24, 2004 from http://epaa.asu.edu/epaa/v10n37

Leef, G. C. (2003). *National Board certification: Is North Carolina getting its money's worth?* Raleigh, NC: John Locke Foundation Inc.

Margolis, J. (2004, February 19). A response to the National Board hoax. *Teachers College Record.* Retrieved June 30, 2004 from http://www.tcrecord.org/PrintContent.asp?ContentID=11277

Moss P. A., Sutherland, L. M., Haniford, L., Miller, R., Johnson, D., Geist, P.K., Koziol, S.M., Star, J.R., & Pecheone, R. L., (2004, July 20). Interrogating the generalizability of portfolio assessments of beginning teachers: A qualitative study, *Education Policy Analysis Archives, 12*(32). Retrieved August 22, 2004 from http://epaa.asu.edu/epaa/v12n32/.

National Board for Professional Teaching Standards (no date). *About NBPTS/History & facts. Looking back at our roots: Early days, critical players.* Retrieved June 26, 2004 from http://www.nbpts.org/about/lookingback.cfm

National Board for Professional Teaching Standards, (1989). *Toward high and rigorous standards for the teaching profession: Initial policies and perspectives of the National Board for Professional Teaching Standards.* Detroit: Author.

National Board for Professional Teaching Standards. (2001). *National Board Certification® candidate survey.* Retrieved September 10, 2002, from http://www.nbpts.org/research/archive_3.cfm?id=9

O'Connor, K. A. *Identifying the wants and needs of North Carolina upper elementary (Grades 3-5) National Board and non-National Board teachers for job success and satisfaction.* Ann Arbor, MI: Proquest Information and Learning Company.

Petty, T. M. (2002). *Identifying the wants and needs of North Carolina high school mathematics teachers for job success and satisfaction.* Ann Arbor, MI: Proquest Information and Learning Company.

Petrosky, A. R. (1994). *Schizophrenia, the National Board for Professional Teaching Standards' policies and me.* Retrieved February 9, 2003, from http://www.dil.sched.pitt.edu/people/faculty/Petrosky/Schizophrenia.html

Petrosky, A. R. (1998, July). Insiders and outsiders: Teaching standards, national certification assessment, and professional development. *English in Australia, 122*, Retrieved June 29, 2004 from http://www.aate.org.au/E_in_A/July98/982petro.html

Podgursky, M. (2001). Defrocking the National Board. *Education Next, 1* (2), 79-82.

Podgursky, M. (2001, April 11). Should states subsidize national certification? *Education Week on the WEB*, Retrieved June 23, 2004 from http://www.edweek.org/ew/ewstory.cfm?slug=30podgursky.h20

Podgursky, M. (2001, Fall). *Education Matters rejoinders to Wenglinsky and Bond.* Retrieved June 30, 2004 from http://www.educationnext.org/unabridged/20013/letter-pod.html

Pool, J. E., Ellett, C. E., Schiavone, S. & Carey-Lewis, C. (2001). How valid are the National Board for Professional Teaching Standards assessments for predicting the quality of actual classroom teaching and learning? Results of six mini case studies. *Journal of Personnel Evaluation in Education, 15* (1), 31-48.

Ralph, D. E. (2003). *National Board Certified Teachers' views of the certification process and its effect on the professional school culture.* Ann Arbor, MI: Proquest Information and Learning Company.

Rapp, D. (2001). Ohio teachers give tests an 'F.' *Rethinking Schools, 15* (4). Retrieved July 1, 2004 from:http://www.rethinkingschools.org/archive/15_04/Ohio154.shtml

Remark Web Survey (Version 2.0). (2002, January). [Computer software]. Paoli, PA: Principia Products.

Sato, M. (2000, April). *The national Board for professional teaching standards; Teacher learning through the assessment process.* Paper presented at the meetings of the American Educational Research Association, New Orleans , LA. Retrieved June 30, 2004 from http://nbrc.stanford.edu/research/learning.pdf

Sato, M., Hyler, M. E., & Monte-Sano, C. (2002, April). *The National Board certification process and its impact on teacher leadership.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. Retrieved July 2, 2004 from http://nbrc.stanford.edu/research/leadership.pdf

Schön, D. (1983). *The reflective practitioner.* New York: Basic Books.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Education Review, 57*, 1-22.

Shulman, L. S. and Quinlan, K. M. (1995). The comparative psychology of school subjects. In D. C. Berliner and R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 399-422). New York: Macmillan.

Schutz, A. & Moss, P.A., (2004, July 20). Reasonable decisions in portfolio assessment: Evaluating complex evidence of teaching, *Education Policy Analysis Archives, 12*(33). Retrieved August 22, 2004 from http://epaa.asu.edu/epaa/v12n33/.

Smith, M.L, & Glass, G.V. (1987). *Research and evaluation in education and the social sciences.* Needham Heights, MA: Allyn and Bacon.

Stanford Achievement Test-9th *Administrator's interpretive guide.* (2001). San Antonio, TX: Harcourt Assessment.

Stephens, A. D. (2003, September). *The relationship between National Board certification for teachers and student achievement*. Unpublished doctoral dissertation, College of Education, University of South Carolina, Columbia.

Sternberg, R. J. & Horvath, J. A. (1995). A prototype of expert teaching. *Educational Researcher, 24*, 9-17.

Stone, J. E. (2004). National Board for Professional Teaching Standards (NBPTS): Improving education at a snail's pace. Retrieved June 23, 2004 from http://www.education-consumers.com/Urban-Inst1.asp

Stone, J. E. (2002). *The value-added achievement gains of NBPTS-certified teachers in Tennessee*: *A brief report*. Retrieved June 30, 2004, from http://www.education-consumers.com/briefs/stoneNBPTS.shtm

Taylor, G. (2000). *Teacher change and the National Board for Professional Teaching Standards: A case study of eleven Colorado teachers.* Unpublished doctoral dissertation, College of Education, University of Colorado, Boulder.

Thirunarayanan, M. O. (2004, February 10). National Board certification for teachers: A billion dollar hoax. *Teachers College Record.* Retrieved June 24, 2004 from http://www.tcrecord.org/Content.asp?ContentID=11266

Turchi, L. B. (1996). *Teacher classroom inquiry: Activities, thinking, reflective stances and collaboration embedded in the assessment of accomplished teaching by the National Board for Professional Teaching Standards.* Unpublished doctoral dissertation, College of Education, Appalachian State University, Boone, North Carolina.

U. S. Department of Education (2003). *Meeting the highly qualified teacher challenge*. Washington, DC: Author. Retrieved June 26, 2004 from http://www.ed.gov/about/reports/annual/teachprep/2003title-ii-report.pdf

United States Commission on Excellence in Education. (1983, April). *A nation at risk*. Washington, DC: Author. Retrieved January 27, 2002, from http://www.ed.gov/pubs/NatAtRisk/risk.html

Van Driel, J. H., Beijaard, D. & Verloop, N. (2001) Professional development and reform in science teaching: The role of teachers' practical knowledge. *Journal of Research on Science Teaching, 38*, 2, 137-155.

Whitman, B. A. (2002, October). *Professional teachers for quality education: Characteristics of teachers certified by the Board for Professional Teaching Standards*. New Brunswick, NJ: Rutgers, The State University of New Jersey.

About the Authors

Leslie Vandevoort
Kyrene School District
Tempe, AZ
Email: Lvande@kyrene.org

Leslie Green Vandevoort earned a B.A. in Psychology and Education from Muskingum College (Ohio), a M.A. in Educational Psychology from the University of Arizona and an Ed. D. in Educational Administration from Arizona State University. She has practiced school psychology in the suburban Tucson and Phoenix areas for more than 25 years.

Audrey Amrein Beardsley
Arizona State University West
College of Teacher Education and Leadership
PO Box 37100, Phoenix, AZ 85069
Work Phone: 602.543.6374
Email: audrey.beardsley@asu.edu
Homepage: http://members.cox.net/audrey.beardsley/

Audrey Amrein Beardsley is an Assistant Professor at Arizona State University – West Campus. She received her PhD from Arizona State University – Tempe in 2002 from the Division of Educational Leadership and Policy Studies in the College of Education. Her research interests include education policy, research methods, and educational assessments. Her current research focuses on high-stakes testing policies and their effects on academic achievement, schools, and students from racial minority and economically disadvantaged backgrounds. Specifically, her research focuses on how high-stakes testing policies impact multiple indicators of academic achievement and school quality in the states that have attached severe consequences to student, teacher, school, and district performance on tests. Her teaching interests include research methodology, assessment, and education policy.

David C. Berliner
Regents' Professor of Education
College of Education
Arizona State University
Tempe, AZ 85287-2411

Email: berliner@asu.edu

David C. Berliner is Regents' Professor of Education at the College of Education of Arizona State University, in Tempe, AZ. He received his Ph.D. in 1968 from Stanford University in educational psychology, and has hled positions at the University of Massachusetts, WestEd, and the University of Arizona. He has served as president of the American Educational Research Association (AERA), president of the Division of Educational Psychology of the American Psychological Association (APA), and as a Fellow of the Center for Advanced Study in the Behavioral Sciences. Berliner is a member of the National Academy of Education. His publications include *The Manufactured Crisis*,

Addison-Wesley, 1995 (with B.J. Biddle) and the *Handbook of Educational Psychology*, Macmillan, 1996 (Edited with R.C. Calfee). Special awards include the Research into Practice Award and the lifetime achievement award from AERA, the E. L. Thorndike award from APA, and the 2003 Brock international award for educational achievements. His scholarly interests include research on teaching and education policy analysis.

Education Policy Analysis Archives http://**epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@uconn.edu.

Appendix A

Two Comments From Teachers

Teacher: "Students and parents assigned to me, frequently have had personality or power conflicts with the teachers in previous years. My students often have a history of emotionally-based underachievement. Rather than comparing my class average with my colleagues' it might be interesting to compare the academic growth my students achieved in my class with the growth they achieved in previous years."

Teacher: "My friend and colleague forwarded your email to me and I realized that I, too, had not responded to your request. …However, I would like to share these thoughts with you. Below is my response to (NBCT's Name) when she forwarded your email to me: I did not respond to her letter, but I think it is very important that she know how we feel about test scores as a measure of our master teacher status. The NBPTS process in no way changed my teaching practices and I continue to stand strongly against any and all standardized testing. I hope that my credential as a NBCT gives me a respected voice with which to state those views, but it has not affected my students' test scores.

Sorry. One of the reasons I entered into the National Board certification process was because there was no mention whatsoever about norm referenced tests when assessing students' progress. What drew me to the standards was their emphasis on a teacher's own authentic assessment – interview, anecdotal records, direct observation, and analysis and reflection on individual student's work samples. Children are tested far too much, and reveal way too little about what they've actually learned on adult-made tests. Children do not think like adults. I think norm referenced testing is a meaningless waste of money in a system where smaller class sizes and paying teachers a competitive salary would make the most significant difference to children. I also abhor the use of norm-referenced tests to base teacher bonuses ("pay for performance") or label schools. It is unconscionable to base adult's salaries on children's performance. Yes, the quality of the teacher probably does impact on test scores, but so do the children's backgrounds, and therefore, sometimes the teachers who are the best and work the hardest do not get recognized for their efforts because of their students' lower test scores. As you know, norm referenced tests are manipulated scores and there always has to be a "below 50th percentile" so there can be an "above 50th percentile." I am very uninterested in comparing my students to the norming group who took the test years ago. I'll continue to teach as long as I can ignore all the nonsense about testing."

Appendix B

Teacher Survey Results

1. What is your date of birth?

Mean	Range
48 years	32-63 years

2. How many years of teaching experience do you have?

Mean	Median	Range
19.75 years	20 years	10-33 years

3. What is your current position?

Classroom Teacher	Teacher on assignment outside classroom	School Administrator	District Administrator	No longer in education	Other
70.5 %	8.8 %	5.8 %	2.9 %	2.9 %	8.8 *

* faculty at local university's education college (2.9%), gifted resource (2.9%), honors math/reading (2.9%)

If you are a classroom teacher, in what grade(s) do you currently teach?

	Pre-K	K	1	2	3	4	5	6
N	1	1	2	4	5	5	2	4

	K/1/2	1/2	2/3	3/4/5	3/4/5/6
N	2	1	1	1	1

4. From what college or university did you receive your teaching degree?

Public/out of state	ASU/ASU West	Private/out of state	U of AZ	Private/in state
44.1 %	29.4 %	17.6%	5.8 %	2.9 %

In what year did you receive your teaching degree?

Median	Range
1978	1963-1996

5. What is the highest degree level you've achieved?

Bachelor's	Master's	Doctorate
5.8%	88.2 %	5.8%

In what year did you receive that degree?

Median	Range
1992	1964-2002

6. Have you taken coursework beyond that degree?

Yes	No
82.3%	17.6 %

If "yes," about how many units have you earned beyond the degree?

Degree	Bachelor's	Master's	Doctorate
Mean number of units	57.5	30.8	25

In what areas of study? (Multiple responses allowed. Findings reported as percent of total respondents).

Area of study	%	Area of study	%
Elementary Education	21.4	Reading	7.1
English as a Second Language	21.4	Math	7.1
Administration	17.9	Professional development	7.1
Technology	14.3	Science	3.6
Education	14.3	Multiple Intelligences	3.6
Gifted	14.3	National Boards	3.6
Special education	10.7	Leadership	3.6
Instructional methods	7.1	Curriculum	3.6
Child development	7.1	Other	3.6

7. Do you participate in professional growth activities on a regular basis?

Yes	No
97 %	3 %

If "yes," in what types of activities have you participated during the past year? (Multiple responses allowed. Findings reported as percent of total respondents).

	Workshops	Professional journals	Conferences	Professional organizations	College/university classes	Other
%	96.8	90.6	81.2	71.8	62.5	28.1

If "other," please explain:
daily professional growth block at school (3.1%), National Board study (3.1%), district classes (3.1%), career ladder program (3.1%), teach language arts methods class at university (3.1%), district in-service (3.1%), on-going professional development at school (3.1%), curriculum writing projects (3.1%), guitar lessons (3.1%)

8. What endorsements or certificate(s) do you currently hold from the state of Arizona? (Multiple responses allowed. Findings reported as percent of total respondents).

Endorsement	%	Endorsement	%
English as a Second Language	16.3	Bilingual Education	2.3
Principal	14.0	Physical Education	2.3
Elementary Education	11.6	Supervisor	2.3
Gifted	9.3	Guidance Counselor	2.3
Reading Specialist	9.3	Emotional Handicaps	2.3
Learning Disabled	9.3	Early Childhood	2.3
Special Education	7.0	Early Childhood Handicapped	2.3
Middle Grades	4.7	Adult Education	2.3

9. If you are no longer a classroom teacher, please tell us the month and year in which you last taught.

N	Range
8	8/1998 to 5/2002

10. If you are no longer a classroom teacher, was your decision to leave the classroom in any way related to your achievement of Board certification? (N=9)

No	Yes
77.7	22.2 %

11. In what years did you administer the SAT-9? (Multiple responses allowed. Findings reported as percent of total respondents).

Year	1997	1998	1999	2000	2001	2002
%	61.8	55.9	55.9	58.9	58.9	52.9

12. If you administered the SAT-9, did you administer it to your own students?

Yes	No
96 %	4 %*

*"I administered it to my own students but on the years I have multi-age, another teachers did my other grades and I did one only."

13. Do you encourage your students to be present during the week that SAT-9's are administered? (N=27)

Yes	No
96.2%	3.7 %

14. How important would you say the SAT-9 is at your school?

Very Important (4)	Somewhat important (3)	Of little importance (2)	Unimportant (1)
67.8 %	28.5 %	0 %	3.57 %

Mean	Mode	Standard Deviation
3.60	4	0.6852

15. Do you change your curricular focus prior to administering SAT-9s? (N=27)

Yes	No
55.5 %	44.4 %

If "yes," please explain what you do differently?

	Test taking strategies	Review math/ language arts	Cover standards	Various writing activities
Percent	57.1 %	28.5 %	7.1 %	7.1 %

16. Are your students' SAT-9 results used as a measure of your job performance?

No	Yes
62.9 %	37.3 %

If "yes," how much importance do you believe is placed on them?

Very Important (4)	Somewhat important (3)	Of little importance (2)	Unimportant (1)
50 %	50 %	0 %	0 %

Mean	Standard deviation
3.5	0.527046277

17. In your opinion, what is the best way to measure teacher quality? (Multiple responses allowed. Findings reported as percent of total respondents).

Measure	%	Measure	%
Observation	46.9	Student input	12.5
Instruction/lesson delivery	21.9	National Board certification	12.5
Student growth over the year	21.9	Use of teacher-developed rubrics	9.4
Parent input	18.8	Adherence to teacher standards	9.4
Student achievement	18.8	Participation in professional growth activities	9.4
Artifacts, portfolio, etc.	15.6	Contribution to profession	6.2

knowledge of pedagogy/content area (3.1%), self-evaluation (3.1%), administrator-teacher conference (3.1%),

18. In your opinion, what is the best way to measure student achievement? (Multiple responses allowed. Findings reported as percent of total respondents).

67.7 % of the respondents used the word "variety" in their answer.

Measure	%	Measure	%
Observation	28.1	Daily work samples	15.6
Yearly growth	28.1	TMA	12.5
Portfolios	25.0	Cumulative projects	6.3
Norm-referenced tests	21.9	Measures of mastery	6.3
Authentic/performance assessment	21.9	Student self-assessment	6.3
Criterion-referenced test	18.8	Other *	34.3

* tests (3.1%), teacher-student interactions (3.1%), running records (3.1%), Bloom's taxonomy (3.1%), products (3.1%), attendance (3.1%), on-going scale (3.1%), individual goals (3.1%), oral presentations (3.1%), district testing (3.1%), videos (3.1%)

19. How is student achievement typically measured in your classroom? (Multiple responses allowed. Findings reported as percent of total respondents).

	Work samples	Teacher-made assessments	Criterion-referenced tests	Norm-referenced tests	Other
%	93.8	93.8	65.6	56.2	65.6

If "other," please list:
Observation (12.5%), performance-based assessments (12.5%), presentations (6.3%), running records (6.3%), dialogue (6.3%), projects (3.1%), peer review (3.1%), self review (3.1%), cooperative learning (3.1%), teamwork (3.1%), personal best (3.1%), rubrics (3.1%)

20. In your classroom, which type of assessment would you say is used most? Please rate each measure on a scale from 1 to 5, with 1 being the most frequently used and 5 being the least frequently used.

	Work samples	Teacher-made assessments	Other	Criterion-referenced tests	Norm-referenced tests
Mean	1.54	2.39	3.03	3.54	4.15
Rank	1	2	3	4	5
Standard dev.	1.033	1.028	1.34	0.904	1.11

21. Do you know the criteria or process used to determine which students are assigned to your classroom?

Yes	No
90 %	10 %

If "yes," please explain: (Multiple responses allowed. Findings reported as percent of total respondents).

Criteria/process	%	Criteria/process	%
High/medium/low academics	48.1	Special needs	18.5
Parent request/input	48.1	Teaching/learning style	14.8
Previous teacher input	37.0	Teacher/student personalities	11.1
High/medium/low behavior	33.3	ELL status	11.1
Gender	25.9	Race	11.1
Principal input	22.2	Teacher recommendation	11.1

22. Do you agree that this is the best way to assign students to classrooms?

Strongly disagree (6)	Disagree (5)	Mildly disagree (4)	Mildly agree (3)	Agree (2)	Strongly Agree (1)
30.7%	0 %	11.5 %	15.3 %	30.7 %	11.5 %

Mean	Mode	Standard deviation
3.5	2, 4	1.88

23. Briefly, how would you describe your teaching style? (Multiple responses allowed. Findings reported as percent of total respondents).

Teaching Style	%	Teaching Style	%	Teaching Style	%
Hands on	31.3	Active involvement	12.5	Real world based	6.3
Student-centered	28.1	Project-based	12.5	Whole class instruction	6.3
Teach to learning styles	25.0	Parent involvement	12.5	Resource/facilitator	6.3
Structured/ordered	21.9	Experiential	9.4	Guide	6.3
Eclectic	15.6	Fun	9.4	Traditional	6.3
Individualized	15.6	Developmentally-based	6.3	Standards-based	6.3
Use of cooperative learning	15.6	Use of centers	6.3	Individual responses*	74.4
Discovery learning	12.5	Innovative/creative	6.3		

*****Provide choices (3.1%), empowering, teacher/colleague planned, emphasis on writing, multi-sensory, use of class meetings, student paired learning, back to basics, model/assist, use of technology, relaxed, independent, based on multiple intelligences, direct instruction, thematic instruction, open, use of trade books, holistic, flexible, outcome-based, pragmatic, use of demonstration lessons, mastery learning, spiral review

24. What were your reasons for obtaining Board certification? (Multiple responses allowed. Findings reported as percent of total respondents).

Reason	%	Reason	%
Personal/professional/intellectual challenge	37.5	Enjoyment of learning	6.3
Professional growth experience	31.3	Provided feedback from others	6.3
Validation of teaching practice	25.0	Allowed for comparison to others/standards	6.3
To align skills to standards	21.9	Intrigued by the process	6.3
Personal self-development	15.6	Motivated by others	6.3
Desire to reflect	12.5	Individual responses*	37.2
To improve effectiveness as a teacher	12.5		

*****Benchmark for teacher professionalism (3.1%), enjoyment of working with other candidates, placed high expectations on self, demanded own personal best, personal fulfillment, platform by which to unify voice and substantiate purpose of teaching, quantifies work, refresh skills, tired of hearing what teachers couldn't do, to prove the principal wrong, uplift the profession, to gain knowledge of what else is going on in the profession

25. Did you receive any type of support as you went through the certification process?

Yes	No
96.8 %	3.1 %

If "yes," what type of assistance did you receive? (Multiple responses allowed. Findings reported as percent of total respondents).

	Financial	University-based	Release time	Mentoring	Other*
%	100	80.6	77.4	74.2	22.4

* support of family (6.4%), support of colleagues/principal (3.2%), support of friend (3.2%), district-based editing/support team (3.2%), help with reading/preparing entries (3.2%), support from another candidate (3.2%)

26. In what ways do you think the Board certification process has made you a better teacher? (Multiple responses allowed. Findings reported as percent of total respondents).

	More reflective	Improved student achievement	More analytical	Standards-based	New teaching approaches	Greater confidence	Greater focus
%	68.8	28.1	15.6	12.6	12.6	9.4	9.4

27. Is there anything else you would like to share regarding your Board certification experience? (Multiple responses allowed. Findings reported as percent of total respondents).

Comment	%
Worthwhile, rewarding experience	38.9
Excellent professional growth activity	38.9
Support, encouragement, help from others was important	27.8
Improved, honed teaching skills	22.2
Made lasting friendships	11.1
Individual responses*	12.6

* made me a better administrator (.6%), humbling experience, motivated students, improved classroom atmosphere, reinforced "best practices," helped to have had career ladder experience, not prepared for assessment center activities, should have an element of teacher observation – not just writing, renewed excitement for the field, excellent program, need to encourage others to do so, became more reflective, bonded with students and parents, improved student achievement/learning, time consuming/ demanding, portfolio was a good activity, validated convictions, assessment center activities were not representative and did not allow enough time, excellent self-improvement activity, provides teachers with a national platform, allows teachers to monitor themselves

APPENDIX C

PRINCIPAL SURVEY RESULTS

1. Please provide the name of your school and the district in which it is located (school/district).

Principals	Schools	Districts
N=24	24	14

2. How long have you been a principal at this location?

	Mean	Range	Mode
Years	4.75	1-14	3

3. What is the name of your NBCT? (Only the N is provided due to confidentiality).

N
35

4. Is the NBCT currently teaching at your school?

Yes	No
91.4%	8.6%

If "no, " please explain:
District teacher mentor program, leave of absence – family reasons, medical leave

5. How would you describe the manner in which children are assigned to classrooms at your school? Please check all that apply. (Multiple responses allowed. Findings reported as percent of total respondents).

Homogeneous/ tracked	Heterogeneous/ stratified	Random assignment	Parent request	Other
80%	40%	22.9%	0%	34.3%

If "other," please describe.
some parent requests are honored (11.4%), teacher collaboration/team (11.4%), teaching and learning match (8.6%), all day K is a tuition-based program therefore parents' choice (2.9%)

6. Do classrooms at a particular grade level in your school have relatively equal numbers of students?

Yes	No
97.1%	2.9%

7. How are special needs students assigned to classrooms? Please check all that apply. (Multiple responses allowed. Findings reported as percent of total respondents).

Method of placement	%
Team decision	94.3
Severity of needs is weighted	80.0
Principal decision	74.3
Teacher(s) recommendation	65.8
Equal numbers per classroom	60.0
Parent request for teacher	51.4
Other	8.8

If "other," please explain.
no set placement procedure – consider individual children and their needs (2.9%), large special education population and inclusion is a big part of the program (2.9%), an "inclusion classroom" is available at certain grade levels (2.9%)

8. In your district, how are teachers "assigned" to schools? (Multiple responses allowed. Findings reported as percent of total respondents).

	Committee decision	Applicant choice	HR decision	Superintendent's decision	Other
%	45.7	31.4	8.6	2.9	42.9

If "other," please explain.
draft process (14.3%), district screening and principal decision (14.3%), principal decision (11.4%), principal/applicant decision (2.9%)

9. In your school, how are teachers assigned to grade levels? (Multiple responses allowed. Findings reported as percent of total respondents).

	Principal decision	Teacher decision	Team decision	Grade level decision	Other
%	88.6%	28.6	20	5.7	5.7

If "other," please explain.

input from grade level team/principal decision (5.7%)

10. Are parents in your school allowed to request a specific teacher for their child?

Yes	No	Other
37.1 %	28.6 %	34.3%

If "other," please explain. (Some of the respondents who answered with either "Yes" or "No" also completed this item).
not allowed to request a specific teacher but allowed input on environment/learning style, etc. (14.3%), parent input "considered" (14.3%), grades 1-5 only (5.7%), specific occasions/based on needs (2.9%), parents are allowed to request one of three possible teachers (2.9%)

11. If parents are allowed to request specific teachers at your school, please describe the frequency with which your NBCT is requested by parents.

Requested more often than other teachers	34.3 %
Requested about as often as other teachers	17.1 %
Requested less often than other teachers	0.0 %
No opinion/Haven't noticed/Not applicable	37.1 %
No response	11.4%

12. If the NBCT is requested more often than his or her peers, to what do you attribute this?

Prior Reputation	Outstanding teacher	Student achievement	Provides quality education	Individual responses
33.3%	9.5%	9.5%	9.5%	19%*

*experienced (4.8%), personality (4.8%), only person in all-day kindergarten (4.8%), good teacher (4.8%)

13. In comparison to his/her grade level teaching peers, would you say that your NBCT has historically been assigned:

More high needs students	14.7 %
The same number of high needs students	70.6 %
Fewer high needs students	0.0 %
Don't know/No opinion/NA	14.7 %

14. When did you supervise your NBCT? (Please check all that apply).

Prior to the Board certification process	42.9%
During the certification process	60.0%

After the certification process	88.6%

Prior/during/after	35.3 %
After only	35.3 %
During/after	17.6 %
Prior/During	5.9 %
Prior/after	2.9 %
During only	2.9 %
Prior only	0.0 %

15. Have you observed any changes in your NBCT's teaching that you might attribute to his/her Board certification experience?

Yes	No	No response
55.9%	23.5%	20.1%

Please explain.

More reflective	29.4%
Assumes more of a leadership role	11.8%
Eager to try new things/Increased risk-taking	8.8%
Affirmed/verified beliefs about practice	5.9%
More confident	5.9%
More analytical	5.9%
Improved instructional delivery	5.9%
Pays greater attention to weaknesses	5.9%
Changes practice to suit student needs	5.9%

goal driven (2.9%), new attitude – "if it doesn't work, change it" (2.9%), utilizes academic standards (2.9%), continues to grow and collaborate (2.9%), extremely motivated to student achievement (2.9%)

16. Compared to all of the teachers you have supervised, would you say that your NBCT is:

the best teacher?	one of the best teachers?	an average teacher?	one of the poorest teachers?	the poorest teacher?
2.9 %	85.3 %	8.8 %	2.9 %	0 %

17. If you rated your NBCT as "the best" or "one of the best" teachers, what are the characteristics that make the NBCT stand out from the rest? (Multiple responses allowed. Findings reported as percent of total respondents).

Teacher Characteristics			
Collaborative	12.5%	Personable	4.2%
Organized	9.7%	Self-reflective	4.2%
Dedicated	8.3%	Initiating	4.2%
Professional/ethical	6.9%	Methodical	4.2%
Motivating/challenging	5.6%	Energetic/enthusiastic	2.8%
Focused/determined	5.6%	Efficient	2.8%
Communicative	5.6%	Driven	2.8%
Leader	5.6%	Prepared	2.8%

creative (1.4%), caring (1.4%), intelligent (1.4%), problem solver (1.4%), mentor/role model (1.4%), positive attitude (1.4%), team player (1.4%), analytical (1.4%), effective (1.4%)

Classroom/Instruction	
Knowledge/expertise of curriculum/instruction/content	39.0%
Uses variety of methods	10.2%
Meets student needs	8.5%
Classroom/time management	6.8%
Tries new things	5.1%
Data driven decisions	5.1%
Hands-on approach	3.4%
Integrates technology into classroom	3.4%
Uses direct instruction	3.4%
Uses best practices	3.4%
Assesses student learning	3.4%

structured classroom (1.7%), experiential/applied (1.7%), teaches to higher order thinking skills (1.7%), co-operative learning environment (1.7%), positive class environment (1.7%)

Other	
Makes gains in student achievement	17.6%
Positive parents/aide relationships	17.6%
Reading Recovery trained	11.8%
Takes on extra responsibilities	11.8%
Sets high standards for students	11.8%

all children learn (5.9%), student oriented (5.9%), establishes rapport with children (5.9%), contributes to school (5.9%), puts in long hours (5.9%)

18. Please rate your NBCT on the following:

	Excellent (6)	**Very Good (5)**	**Good (4)**	**Fair (3)**	**Poor (2)**	**Unacceptable (1)**

Relationships with colleagues	41.2%	47.1%	11.8%	0%	0%	0%
Relationships with parents	52.9%	38.2%	8.8%	0%	0%	0%
Relationships with students	67.6%	17.6%	14.7%	0%	0%	0%
Classroom management skills	76.7%	6.7%	13.3%	0%	3.3%	0%
Use of instructional strategies	76.5%	14.7%	0%	2.9%	5.9%	0%
Skills at assessing student learning	70.6%	20.6%	0%	2.9%	5.9%	0%
Planning/goal setting	67.6%	17.6%	8.8%	2.9%	2.9%	0%
Mean	64.7%	23.2%	8.2%	1.2%	2.6%	0%

Relationships with colleagues	
Mean	5.3
Median	5
Mode	5
Stand. Dev.	0.7022
Range	2
Minimum	4
Maximum	6
Sum	159

Relationships with parents	
Mean	5.4
Median	5.5
Mode	6
Stand. Dev	0.6747
Range	2
Minimum	4
Maximum	6
Sum	162

Relationships with students	
Mean	5.47
Median	6
Mode	6
Stand. Dev	0.7761
Range	2
Minimum	4
Maximum	6
Sum	164

Classroom management	
Mean	5.53
Median	6
Mode	6
Stand. Dev	0.9732
Range	4
Minimum	2
Maximum	6
Sum	166

Use of instructional strategies	
Mean	5.5
Median	6
Mode	6
Stand. Dev	1.1371
Range	4
Minimum	2
Maximum	6
Sum	165

Skills at assessing student learning	
Mean	5.43
Median	6
Mode	6
Stand. Dev	1.1351
Range	4
Minimum	2
Maximum	6
Sum	163

Planning and goal setting	
Mean	5.4
Median	6
Mode	6
Stand. Dev	1.0372
Range	4
Minimum	2
Maximum	6
Sum	162

Total teacher ratings	
Mean	38.0333
Median	40
Mode	42
Stand. Dev	5.5179
Range	22
Minimum	20
Maximum	42
Sum	1141

19. Did you observe any change in your NBCT's SAT-9 scores after he/she became Board certified?

No response	No	Yes
45.2 %	19.4%	35.5%

Please explain.
slight increase in students scores overall, her higher scores on the SAT-9 mean NCE column were higher than the previous year, she has always had excellent scores but more of her "low" students made significant progress

20. As compared to other teachers in your school, how would you describe your NBCT's level of involvement in professional growth activities?

More involved	About the same	Less involved	Don't know
76.5 %	20.6 %	0 %	2.9 %

21. Has the level of his/her involvement in professional growth activities changed since the Board certification experience?

There has been an increase in the level of professional growth activities	41.2 %
There has been no change in the level of professional growth activities	17.6 %
There has been a decrease in the level of professional growth activities	41.2 %
Don't know	0 %

22. If you were supervising the NBCT during the year (s)he went through the Board certification process, about how much time would you say you spent assisting the candidate to achieve National Board certification?

Very little, small amount, etc.	47.1%
10-15 hours	11.8%
10 plus hours	5.9%
5 hours	5.9%
4 hours	5.9%
2 hours per week	5.9%

"I'm not sure I can quantify the amount of time I spent in supporting the NBCT during this process. I can say, however, that our conversation was definitely more intense (e.g. - watching videos of her teaching and then debriefing these with her." (5.9%)

"…My assistant principal spent a great deal of time helping her with taping and reviewing her documents." (5.9%)

"Difficult to pinpoint. I will say that I was an avid supporter!" "Unfortunately not as much as I should have .I was a first year principal, and was not familiar with the job, staff or teachers. The NBCT is very self-motivated self started and would ask me for my assistance when needed. In the future I will be more available and involved with any of my teachers that choose to earn NBCT." (5.9%)

23. Do you believe the amount of time you spent was worthwhile? (Percentages are reported only for those principals who responded to this item).

Yes	Not Sure	No
90%	10 %	0 %

24. To what degree do you believe each of these variables influences a student's SAT-9 scores?

Variable	Very influential (5)	Somewhat influential (4)	Not very influential (3)	Not at all influential (2)	Don't know (1)
Student's socio-economic level	50%	46.7%	3.3%	0%	0%
Teacher quality during the given year	90%	6.7%	3.3%	0%	0%
Parent involvement in student's education	79.3%	20.7%	0%	0%	0%
Teacher quality across the years	90%	10%	0%	0%	0%
Parents' education level	26.7%	70%	3.3%	0%	0%

Student ability	73.3%	20%	6.7%	0%	0%
Test awareness/ preparation	30%	46.7%	23.3%	0%	0%
Mean	62.8%	31.5%	5.7%	0%	0%

Student SES	
Mean	4.47
Standard Error	0.1043
Median	4.5
Mode	5
Standard Dev.	0.5713
Range	2
Minimum	3
Maximum	5
Sum	134

Teacher quality	
Mean	4.87
Standard Error	0.0793
Median	5
Mode	5
Standard Dev.	0.4342
Range	2
Minimum	3
Maximum	5
Sum	146

Parent involvement	
Mean	4.79
Standard Error	0.0766
Median	5
Mode	5
Standard Dev.	0.4123
Range	1
Minimum	4
Maximum	5
Sum	139

Teacher quality over time	
Mean	4.9
Standard Error	0.0557
Median	5
Mode	5
Stand. Dev.	0.3051
Range	1
Minimum	4
Maximum	5
Sum	147

Parent's educational level	
Mean	4.23
Standard Error	0.0920
Median	4
Mode	4
Stand. Dev.	0.5040
Range	2
Minimum	3
Maximum	5
Sum	127

Student ability	
Mean	4.67
Standard Error	0.1107
Median	5
Mode	5
Stand. Dev.	0.6065
Range	2
Minimum	3
Maximum	5
Sum	140

Test awareness	
Mean	4.07
Standard Error	0.1350
Median	4
Mode	4
Stand. Dev.	0.7397
Range	2
Minimum	3
Maximum	5
Sum	122

25. In your opinion, what is the best way to measure teacher quality?

Direct observation of teaching	29.8%
Student growth/achievement	28.1%
Standards based rubrics	10.5%
Feedback/dialogue	7.0%
Parent feedback	3.5%
On-going training/professional growth	3.5%
Student time on-task/engaged	3.5%

attitude (1.8%), student's ability to apply knowledge (1.8%), portfolios, ability to collaborate with peers (1.8%), ability to deal with parent/colleague issues (1.8%), commitment to students/profession (1.8%), student motivation levels (1.8%), on-going assessment of student performance (1.8%)

26. In your opinion, what is the best way to measure student achievement?

Variety of ways/multifaceted	56.9%*****	Standards-based	7.8%
On-going/growth over time	23.5%	Student attitude toward learning	3.9%
Performance-based	7.8%		

(Multiple responses allowed. Findings reported as percent of total respondents).

***Variety of ways/multifaceted assessment – including:**			
CRT's/NRT's	31.0%	1:1 responses to teacher questions	3.4%
Observation(s)	20.7%	Diagnostic reading assessments	3.4%
Teacher-made assessments	13.8%	Application of learning	3.4%
Running records	6.9%	Pivot tables	3.4%
Portfolio assessment	6.9%	Individual assessment of student by teacher	3.4%
Narratives	3.4%		

ability to apply new learning (3.4%), student attendance rates(3.4%), ability to monitor/adjust (3.4%), student's social/emotional growth (3.4%), student on-task/engagement time (3.4%), student-teacher dialogue (3.4%)

27. Do you believe that standardized test scores are a good measure of teacher quality*?*

No	**Yes**	**No opinion**
60.6 %	39.4%	0 %

Explain: (Multiple responses allowed. Findings reported as percent of total respondents).

One variable among many	40.0%
Biased	22.9%
Prefer yearly growth measures/CRT's	11.4%
Test doesn't match instruction	8.6%
One time measure	5.7%
Overused	2.9%

28. Do you believe the National Board for Professional Teaching Standards is contributing to the improvement of teacher quality?

Yes	No	Don't know/No opinion
91.2%	0 %	8.8 %

29. Do you believe the National Board for Professional Teaching Standards is contributing to improvements in student achievement?

Yes	No	Don't know/No opinion
70.6%	0 %	29.4%

30. Please feel free to provide any additional comments here.

"I believe that the best pathway to increasing student achievement lies in having teachers who are more knowledgeable about 'content' and their professional practice. I also believe that the NCTB Certification has the potential of increasing other teachers' interest in pursuing rigorous professional development that will help us achieve the gains we want in academic and social / behavioral achievement."

"I believe it is more important if the individual has a mindset to enhance themselves professionally. I think even with teachers that are not exemplary, that the NBCT process is a help and a great tool to encourage reflection and attention to national standards that are appropriate and benefit children's learning. Some teachers are more willing to go this route than participate in district sponsored training or university coursework. I think some teachers do this for the wrong reason, that they want the label. Often these teachers are overwhelmed by the amount of work it takes to accomplish this label, but some persevere and succeed not because they are exemplary teachers but because they are willing to persevere. I question the certification process because I believe that there are some teachers that are selected that do not seem to deserve this certification professionally compared to other teachers that are in the NBCT ranks or even some that have not gone through the process but are truly exceptional."

"I would love to see a National Board Certified Principalship program. Keep looking for ways to help cover costs to the teachers."

"I believe there are flaws in the board certification process. I had a teacher who is qualified, effective and an expert in her field and did not get the NCTB recognition after a year's work. As a result, the teacher became very critical of herself and lost self-esteem qualities. I think the key to effective teaching is teachers having support and staff development that is meaningful. That means staff development is not canned, rather it is tailored to meet individual teacher needs. Also, it does not mean that programs are purchased and teachers are required to teach them. We need to hire excellent teachers, nurture them, support them and they will exceed our expectations. How's that for 2 cents worth? Thanks for this opportunity."

"I think the program is an OUTSTANDING vehicle for improvement to the entire educational system. The learning process is rigorous and focused on Best Practice that benefits all learners. It would be so wonderful is this program could be aligned to higher education, for example to receiving credit toward a Doctorate Degree. This would encourage more teachers to pursue the certification due to the benefit of the learning process and financially on salary schedules. Please keep me posted as to any results of this survey or anyway ____ Elementary School can assist this excellent program. Good luck!"

"I too am a Nationally Board Certified teacher and I believe this is beneficial as I evaluate teacher."

"My NBCT's National Board certification has been a very positive achievement for our School, District, and County."

"My NBCT is what I believe NCBT is striving to produce. I suspect that she was already very good before the process, so I'm not sure how much it influenced her. She is a strong advocate for it and I believe that the process and the increased reflection about standards is a good thing for all teachers. I wonder though if it puts more pressure on an administrator if he/she is faced with a NCBT candidate that shows marginal skills, but has the NCBT recognition that could interfere with any efforts to help the teacher improve since they have already been recognized as exceptional."

"My NBCT is the only NBCT I have worked with and all my information is based on her performance."

Appendix D
SPSS Output: Univariate Analysis Of Variance

1999-2000 - READING

Grade = 3

Between-Subjects Factors[a]

		N
NBCT00	0	14506
	1	113

a. Grade = 3

Descriptive Statistics[a]

Dependent Variable: RGN9900

NBCT00	Mean	Std. Deviation	N
0	32.2666	26.67146	14506
1	36.9204	26.68137	113
Total	32.3026	26.67374	14619

a. Grade = 3

Tests of Between-Subjects Effects[b]

Dependent Variable: RGN9900

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	593654.016[a]	2	296827.008	442.386	.000
Intercept	960192.461	1	960192.461	1431.054	.000
READSS99	591225.694	1	591225.694	881.152	.000
NBCT00	1523.440	1	1523.440	2.271	.132
Error	9806881.193	14616	670.969		
Total	25654866.0	14619			
Corrected Total	10400535.2	14618			

a. R Squared = .057 (Adjusted R Squared = .057)

b. Grade = 3

Estimated Marginal Means

NBCT00[b]

Dependent Variable: RGN9900

NBCT00	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	32.274[a]	.215	31.853	32.696
1	35.960[a]	2.437	31.184	40.737

a. Covariates appearing in the model are evaluated at the following values: R scale score = 597.36.

b. Grade = 3

Grade = 4

Between-Subjects Factors[a]

		N
NBCT00	0	15487
	1	184

a. Grade = 4

Descriptive Statistics[a]

Dependent Variable: RGN9900

NBCT00	Mean	Std. Deviation	N
0	28.3773	24.12391	15487
1	27.4620	22.17340	184
Total	28.3665	24.10148	15671

a. Grade = 4

Tests of Between-Subjects Effects[b]

Dependent Variable: RGN9900

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1033326.328[a]	2	516663.164	1003.222	.000
Intercept	1419696.272	1	1419696.272	2756.671	.000
READSS99	1033173.979	1	1033173.979	2006.148	.000
NBCT00	1.203	1	1.203	.002	.961
Error	8069082.284	15668	515.004		
Total	21712242.0	15671			
Corrected Total	9102408.612	15670			

a. R Squared = .114 (Adjusted R Squared = .113)

b. Grade = 4

Estimated Marginal Means

NBCT00[b]

Dependent Variable: RGN9900

NBCT00	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	28.366[a]	.182	28.008	28.723
1	28.447[a]	1.673	25.167	31.726

a. Covariates appearing in the model are evaluated at the following values: R scale score = 628.46.

b. Grade = 4

Grade = 5

Between-Subjects Factors[a]

		N
NBCT00	0	15550
	1	77

a. Grade = 5

Descriptive Statistics[a]

Dependent Variable: RGN9900

NBCT00	Mean	Std. Deviation	N
0	12.3763	22.65173	15550
1	15.5714	23.23814	77
Total	12.3921	22.65500	15627

a. Grade = 5

Tests of Between-Subjects Effects[b]

Dependent Variable: RGN9900

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1753551.664[a]	2	876775.832	2186.036	.000
Intercept	1720491.320	1	1720491.320	4289.644	.000
READSS99	1752769.473	1	1752769.473	4370.122	.000
NBCT00	3777.361	1	3777.361	9.418	.002
Error	6266477.075	15624	401.080		
Total	10419767.0	15627			
Corrected Total	8020028.739	15626			

a. R Squared = .219 (Adjusted R Squared = .219)

b. Grade = 5

Estimated Marginal Means

NBCT00[b]

Dependent Variable: RGN9900

NBCT00	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	12.357[a]	.161	12.043	12.672
1	19.381[a]	2.283	14.906	23.856

a. Covariates appearing in the model are evaluated at the following values: R scale score = 655.92.

b. Grade = 5

Grade = 6

Between-Subjects Factors[a]

		N
NBCT00	0	11752
	1	79

a. Grade = 6

Descriptive Statistics[a]

Dependent Variable: RGN9900

NBCT00	Mean	Std. Deviation	N
0	11.0027	19.90983	11752
1	7.5063	20.82151	79
Total	10.9794	19.91717	11831

a. Grade = 6

Tests of Between-Subjects Effects[b]

Dependent Variable: RGN9900

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	874061.662[a]	2	437030.831	1353.610	.000
Intercept	883590.651	1	883590.651	2736.734	.000
READSS99	873102.353	1	873102.353	2704.249	.000
NBCT00	29.093	1	29.093	.090	.764
Error	3818825.306	11828	322.863		
Total	6119075.000	11831			
Corrected Total	4692886.968	11830			

a. R Squared = .186 (Adjusted R Squared = .186)

b. Grade = 6

Estimated Marginal Means

NBCT00[b]

Dependent Variable: RGN9900

NBCT00	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	10.975[a]	.166	10.650	11.300
1	11.585[a]	2.023	7.619	15.550

a. Covariates appearing in the model are evaluated at the following values: R scale score = 667.55.

b. Grade = 6

1999-2000 - MATH

Grade = 3

Between-Subjects Factors[a]

		N
NBCT00	0	15231
	1	121

a. Grade = 3

Descriptive Statistics[a]

Dependent Variable: MGN9900

NBCT00	Mean	Std. Deviation	N
0	28.5332	28.58077	15231
1	35.5207	24.46497	121
Total	28.5883	28.55665	15352

a. Grade = 3

Tests of Between-Subjects Effects[b]

Dependent Variable: MGN9900

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1195082.476[a]	2	597541.238	809.975	.000
Intercept	1610715.816	1	1610715.816	2183.347	.000
MATHSS99	1189221.243	1	1189221.243	1612.005	.000
NBCT00	4215.654	1	4215.654	5.714	.017
Error	11323383.9	15349	737.728		
Total	25065483.0	15352			
Corrected Total	12518466.4	15351			

a. R Squared = .095 (Adjusted R Squared = .095)

b. Grade = 3

Estimated Marginal Means

NBCT00[b]

Dependent Variable: MGN9900

NBCT00	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	28.542[a]	.220	28.110	28.973
1	34.468[a]	2.469	29.628	39.308

a. Covariates appearing in the model are evaluated at the following values: M scale score = 586.43.

b. Grade = 3

Grade = 4

Between-Subjects Factors[a]

		N
NBCT00	0	15962
	1	186

a. Grade = 4

Descriptive Statistics[a]

Dependent Variable: MGN9900

NBCT00	Mean	Std. Deviation	N
0	32.8877	26.22524	15962
1	28.3978	24.52672	186
Total	32.8360	26.20997	16148

a. Grade = 4

Tests of Between-Subjects Effects[b]

Dependent Variable: MGN9900

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1818823.151[a]	2	909411.575	1583.260	.000
Intercept	2351755.920	1	2351755.920	4094.340	.000
MATHSS99	1815116.859	1	1815116.859	3160.067	.000
NBCT00	1459.512	1	1459.512	2.541	.111
Error	9273557.294	16145	574.392		
Total	28503153.0	16148			
Corrected Total	11092380.4	16147			

a. R Squared = .164 (Adjusted R Squared = .164)

b. Grade = 4

Estimated Marginal Means

NBCT00[b]

Dependent Variable: MGN9900

NBCT00	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	32.868[a]	.190	32.497	33.240
1	30.051[a]	1.758	26.606	33.496

a. Covariates appearing in the model are evaluated at the following values: M scale score = 612.12.

b. Grade = 4

Grade = 5

Between-Subjects Factors[a]

		N
NBCT00	0	16021
	1	82

a. Grade = 5

Descriptive Statistics[a]

Dependent Variable: MGN9900

NBCT00	Mean	Std. Deviation	N
0	24.4158	24.58008	16021
1	34.7561	21.05933	82
Total	24.4685	24.57389	16103

a. Grade = 5

Tests of Between-Subjects Effects[b]

Dependent Variable: MGN9900

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	850735.635[a]	2	425367.818	771.838	.000
Intercept	1107644.265	1	1107644.265	2009.841	.000
MATHSS99	842012.747	1	842012.747	1527.848	.000
NBCT00	11738.567	1	11738.567	21.300	.000
Error	8872878.121	16100	551.110		
Total	19364588.0	16103			
Corrected Total	9723613.756	16102			

a. R Squared = .087 (Adjusted R Squared = .087)

b. Grade = 5

Estimated Marginal Means

NBCT00[b]

Dependent Variable: MGN9900

NBCT00	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	24.407[a]	.185	24.044	24.771
1	36.404[a]	2.593	31.322	41.486

a. Covariates appearing in the model are evaluated at the following values: M scale score = 642.48.

b. Grade = 5

Grade = 6

Between-Subjects Factors[a]

		N
NBCT00	0	11998
	1	79

a. Grade = 6

Descriptive Statistics[a]

Dependent Variable: MGN9900

NBCT00	Mean	Std. Deviation	N
0	21.0604	22.57576	11998
1	20.6203	25.46003	79
Total	21.0575	22.59467	12077

a. Grade = 6

Tests of Between-Subjects Effects[b]

Dependent Variable: MGN9900

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	414121.035[a]	2	207060.517	434.723	.000
Intercept	531760.304	1	531760.304	1116.428	.000
MATHSS99	414105.828	1	414105.828	869.413	.000
NBCT00	458.131	1	458.131	.962	.327
Error	5750905.970	12074	476.305		
Total	11520214.0	12077			
Corrected Total	6165027.005	12076			

a. R Squared = .067 (Adjusted R Squared = .067)

b. Grade = 6

Estimated Marginal Means

NBCT00[b]

Dependent Variable: MGN9900

NBCT00	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	21.042[a]	.199	20.651	21.432
1	23.460[a]	2.457	18.643	28.276

a. Covariates appearing in the model are evaluated at the following values: M scale score = 664.36.

b. Grade = 6

1999-2000 - LANGUAGE

Grade = 3

Between-Subjects Factors[a]

		N
NBCT00	0	15207
	1	123

a. Grade = 3

Descriptive Statistics[a]

Dependent Variable: LGN9900

NBCT00	Mean	Std. Deviation	N
0	29.1972	28.18485	15207
1	30.3496	25.15981	123
Total	29.2065	28.16133	15330

a. Grade = 3

Tests of Between-Subjects Effects[b]

Dependent Variable: LGN9900

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	430380.438[a]	2	215190.219	281.264	.000
Intercept	689156.966	1	689156.966	900.760	.000
LANGSS99	430218.407	1	430218.407	562.315	.000
NBCT00	145.345	1	145.345	.190	.663
Error	11726443.1	15327	765.084		
Total	25233577.0	15330			
Corrected Total	12156823.6	15329			

a. R Squared = .035 (Adjusted R Squared = .035)

b. Grade = 3

Estimated Marginal Means

NBCT00[b]

Dependent Variable: LGN9900

NBCT00	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	29.198[a]	.224	28.758	29.637
1	30.289[a]	2.494	25.401	35.178

a. Covariates appearing in the model are evaluated at the following values: L scale score = 571.54.

b. Grade = 3

Grade = 4

Between-Subjects Factors[a]

		N
NBCT00	0	15955
	1	190

a. Grade = 4

Descriptive Statistics[a]

Dependent Variable: LGN9900

NBCT00	Mean	Std. Deviation	N
0	21.6762	25.93237	15955
1	19.0789	23.18705	190
Total	21.6456	25.90262	16145

a. Grade = 4

Tests of Between-Subjects Effects[b]

Dependent Variable: LGN9900

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3289350.521[a]	2	1644675.261	3519.882	.000
Intercept	3599945.659	1	3599945.659	7704.490	.000
LANGSS99	3288083.903	1	3288083.903	7037.053	.000
NBCT00	1.816	1	1.816	.004	.950
Error	7542397.236	16142	467.253		
Total	18396231.0	16145			
Corrected Total	10831747.8	16144			

a. R Squared = .304 (Adjusted R Squared = .304)

b. Grade = 4

Estimated Marginal Means

NBCT00[b]

Dependent Variable: LGN9900

NBCT00	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	21.647[a]	.171	21.311	21.982
1	21.548[a]	1.568	18.474	24.623

a. Covariates appearing in the model are evaluated at the following values: L scale score = 597.74.

b. Grade = 4

Grade = 5

Between-Subjects Factors[a]

		N
NBCT00	0	15987
	1	81

a. Grade = 5

Descriptive Statistics[a]

Dependent Variable: LGN9900

NBCT00	Mean	Std. Deviation	N
0	11.3084	23.02630	15987
1	13.8765	22.41226	81
Total	11.3213	23.02328	16068

a. Grade = 5

Tests of Between-Subjects Effects[b]

Dependent Variable: LGN9900

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	769092.934[a]	2	384546.467	797.378	.000
Intercept	810028.445	1	810028.445	1679.638	.000
LANGSS99	768561.393	1	768561.393	1593.654	.000
NBCT00	1703.094	1	1703.094	3.531	.060
Error	7747567.081	16065	482.264		
Total	10576133.0	16068			
Corrected Total	8516660.015	16067			

a. R Squared = .090 (Adjusted R Squared = .090)

b. Grade = 5

Estimated Marginal Means

NBCT00[b]

Dependent Variable: LGN9900

NBCT00	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	11.298[a]	.174	10.958	11.639
1	15.896[a]	2.441	11.112	20.680

a. Covariates appearing in the model are evaluated at the following values: L scale score = 619.34.

b. Grade = 5

Grade = 6

Between-Subjects Factors[a]

		N
NBCT00	0	11930
	1	79

a. Grade = 6

Descriptive Statistics[a]

Dependent Variable: LGN9900

NBCT00	Mean	Std. Deviation	N
0	13.0260	21.91115	11930
1	12.3544	27.08677	79
Total	13.0216	21.94786	12009

a. Grade = 6

Tests of Between-Subjects Effects[b]

Dependent Variable: LGN9900

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	822752.155[a]	2	411376.078	995.440	.000
Intercept	868832.863	1	868832.863	2102.386	.000
LANGSS99	822716.762	1	822716.762	1990.795	.000
NBCT00	1017.920	1	1017.920	2.463	.117
Error	4961605.259	12006	413.260		
Total	7820618.000	12009			
Corrected Total	5784357.414	12008			

a. R Squared = .142 (Adjusted R Squared = .142)

b. Grade = 6

Estimated Marginal Means

NBCT00[b]

Dependent Variable: LGN9900

NBCT00	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	12.998[a]	.186	12.633	13.363
1	16.602[a]	2.289	12.115	21.090

a. Covariates appearing in the model are evaluated at the following values: L scale score = 628.44.

b. Grade = 6

2000-2001 - READING

Grade = 3

Between-Subjects Factors[a]

		N
NBCT01	0	15261
	1	169

a. Grade = 3

Descriptive Statistics[a]

Dependent Variable: RGN01

NBCT01	Mean	Std. Deviation	N
0	31.1814	26.47590	15261
1	36.5207	28.21792	169
Total	31.2399	26.50045	15430

a. Grade = 3

Tests of Between-Subjects Effects[b]

Dependent Variable: RGN01

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	772669.691[a]	2	386334.846	592.284	.000
Intercept	1209472.108	1	1209472.108	1854.224	.000
READSS00	767904.646	1	767904.646	1177.263	.000
NBCT01	6437.983	1	6437.983	9.870	.002
Error	10062714.1	15427	652.279		
Total	25894026.0	15430			
Corrected Total	10835383.8	15429			

a. R Squared = .071 (Adjusted R Squared = .071)

b. Grade = 3

Estimated Marginal Means

NBCT01[b]

Dependent Variable: RGN01

NBCT01	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	31.172[a]	.207	30.767	31.577
1	37.379[a]	1.965	33.527	41.230

a. Covariates appearing in the model are evaluated at the following values: R scale score = 599.54.

b. Grade = 3

Grade = 4

Between-Subjects Factors[a]

		N
NBCT01	0	16313
	1	154

a. Grade = 4

Descriptive Statistics[a]

Dependent Variable: RGN01

NBCT01	Mean	Std. Deviation	N
0	29.2084	24.45232	16313
1	27.5000	24.54481	154
Total	29.1924	24.45299	16467

a. Grade = 4

Tests of Between-Subjects Effects[b]

Dependent Variable: RGN01

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1394493.405[a]	2	697246.703	1358.303	.000
Intercept	1809726.440	1	1809726.440	3525.520	.000
READSS00	1394048.128	1	1394048.128	2715.739	.000
NBCT01	113.319	1	113.319	.221	.638
Error	8451329.735	16464	513.322		
Total	23878982.0	16467			
Corrected Total	9845823.140	16466			

a. R Squared = .142 (Adjusted R Squared = .142)

b. Grade = 4

Estimated Marginal Means

NBCT01[b]

Dependent Variable: RGN01

NBCT01	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	29.201[a]	.177	28.853	29.548
1	28.339[a]	1.826	24.760	31.917

a. Covariates appearing in the model are evaluated at the following values: R scale score = 628.90.

b. Grade = 4

Grade = 5

Between-Subjects Factors[a]

		N
NBCT01	0	16682
	1	89

a. Grade = 5

Descriptive Statistics[a]

Dependent Variable: RGN01

NBCT01	Mean	Std. Deviation	N
0	12.6762	22.37154	16682
1	11.0000	19.94538	89
Total	12.6673	22.35916	16771

a. Grade = 5

Tests of Between-Subjects Effects[b]

Dependent Variable: RGN01

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2017651.936[a]	2	1008825.968	2657.152	.000
Intercept	1949815.261	1	1949815.261	5135.630	.000
READSS00	2017403.193	1	2017403.193	5313.650	.000
NBCT01	500.972	1	500.972	1.320	.251
Error	6366211.167	16768	379.664		
Total	11074964.0	16771			
Corrected Total	8383863.103	16770			

a. R Squared = .241 (Adjusted R Squared = .241)

b. Grade = 5

Estimated Marginal Means

NBCT01[b]

Dependent Variable: RGN01

NBCT01	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	12.655[a]	.151	12.359	12.950
1	15.034[a]	2.066	10.985	19.084

a. Covariates appearing in the model are evaluated at the following values: R scale score = 656.19.

b. Grade = 5

Grade = 6

Between-Subjects Factors[a]

		N
NBCT01	0	12642
	1	64

a. Grade = 6

Descriptive Statistics[a]

Dependent Variable: RGN01

NBCT01	Mean	Std. Deviation	N
0	12.3078	19.97998	12642
1	11.5625	18.65976	64
Total	12.3040	19.97293	12706

a. Grade = 6

Tests of Between-Subjects Effects[b]

Dependent Variable: RGN01

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1022296.701[a]	2	511148.351	1604.841	.000
Intercept	1027198.880	1	1027198.880	3225.074	.000
READSS00	1022261.332	1	1022261.332	3209.572	.000
NBCT01	1398.884	1	1398.884	4.392	.036
Error	4045955.832	12703	318.504		
Total	6991803.000	12706			
Corrected Total	5068252.534	12705			

a. R Squared = .202 (Adjusted R Squared = .202)

b. Grade = 6

Estimated Marginal Means

NBCT01[b]

Dependent Variable: RGN01

NBCT01	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	12.280[a]	.159	11.969	12.592
1	16.972[a]	2.233	12.595	21.349

a. Covariates appearing in the model are evaluated at the following values: R scale score = 665.93.

b. Grade = 6

2000-2001 - MATH

Grade = 3

Between-Subjects Factors[a]

		N
NBCT01	0	16116
	1	173

a. Grade = 3

Descriptive Statistics[a]

Dependent Variable: MGN01

NBCT01	Mean	Std. Deviation	N
0	25.0541	28.84421	16116
1	23.7572	26.01272	173
Total	25.0403	28.81519	16289

a. Grade = 3

Tests of Between-Subjects Effects[b]

Dependent Variable: MGN01

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1306252.529[a]	2	653126.264	870.591	.000
Intercept	1649107.448	1	1649107.448	2198.195	.000
MATHSS00	1305964.650	1	1305964.650	1740.799	.000
NBCT01	.107	1	.107	.000	.990
Error	12217918.0	16286	750.210		
Total	23737672.0	16289			
Corrected Total	13524170.5	16288			

a. R Squared = .097 (Adjusted R Squared = .096)

b. Grade = 3

Estimated Marginal Means

NBCT01[b]

Dependent Variable: MGN01

NBCT01	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	25.040[a]	.216	24.617	25.463
1	25.065[a]	2.083	20.983	29.147

a. Covariates appearing in the model are evaluated at the following values: M scale score = 590.73.

b. Grade = 3

Grade = 4

Between-Subjects Factors[a]

		N
NBCT01	0	16786
	1	154

a. Grade = 4

Descriptive Statistics[a]

Dependent Variable: MGN01

NBCT01	Mean	Std. Deviation	N
0	30.4486	25.81665	16786
1	26.2857	24.26776	154
Total	30.4107	25.80534	16940

a. Grade = 4

Tests of Between-Subjects Effects[b]

Dependent Variable: MGN01

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1900238.018[a]	2	950119.009	1715.637	.000
Intercept	2345269.940	1	2345269.940	4234.871	.000
MATHSS00	1897593.534	1	1897593.534	3426.498	.000
NBCT01	810.625	1	810.625	1.464	.226
Error	9379704.026	16937	553.800		
Total	26946280.0	16940			
Corrected Total	11279942.0	16939			

a. R Squared = .168 (Adjusted R Squared = .168)

b. Grade = 4

Estimated Marginal Means

NBCT01[b]

Dependent Variable: MGN01

NBCT01	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	30.432[a]	.182	30.076	30.788
1	28.127[a]	1.897	24.409	31.844

a. Covariates appearing in the model are evaluated at the following values: M scale score = 614.83.

b. Grade = 4

Grade = 5

Between-Subjects Factors[a]

		N
NBCT01	0	17294
	1	84

a. Grade = 5

Descriptive Statistics[a]

Dependent Variable: MGN01

NBCT01	Mean	Std. Deviation	N
0	23.9906	24.06381	17294
1	24.9643	23.98667	84
Total	23.9953	24.06285	17378

a. Grade = 5

Tests of Between-Subjects Effects[b]

Dependent Variable: MGN01

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	969917.127[a]	2	484958.563	926.794	.000
Intercept	1169505.087	1	1169505.087	2235.017	.000
MATHSS00	969837.870	1	969837.870	1853.437	.000
NBCT01	1091.717	1	1091.717	2.086	.149
Error	9091722.487	17375	523.265		
Total	20067432.0	17378			
Corrected Total	10061639.6	17377			

a. R Squared = .096 (Adjusted R Squared = .096)

b. Grade = 5

Estimated Marginal Means

NBCT01[b]

Dependent Variable: MGN01

NBCT01	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	23.978[a]	.174	23.637	24.319
1	27.593[a]	2.497	22.699	32.486

a. Covariates appearing in the model are evaluated at the following values: M scale score = 644.32.

b. Grade = 5

Grade = 6

Between-Subjects Factors[a]

		N
NBCT01	0	12931
	1	63

a. Grade = 6

Descriptive Statistics[a]

Dependent Variable: MGN01

NBCT01	Mean	Std. Deviation	N
0	20.7017	22.58337	12931
1	16.5079	23.17665	63
Total	20.6814	22.58725	12994

a. Grade = 6

Tests of Between-Subjects Effects[b]

Dependent Variable: MGN01

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	463930.912[a]	2	231965.456	488.811	.000
Intercept	549353.408	1	549353.408	1157.629	.000
MATHSS00	462828.249	1	462828.249	975.298	.000
NBCT01	5.282	1	5.282	.011	.916
Error	6164888.049	12991	474.551		
Total	12186612.0	12994			
Corrected Total	6628818.960	12993			

a. R Squared = .070 (Adjusted R Squared = .070)

b. Grade = 6

Estimated Marginal Means

NBCT01[b]

Dependent Variable: MGN01

NBCT01	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	20.683[a]	.192	20.307	21.058
1	20.392[a]	2.747	15.007	25.777

a. Covariates appearing in the model are evaluated at the following values: M scale score = 663.95.

b. Grade = 6

2000-2001 - LANGUAGE

Grade = 3

Between-Subjects Factors[a]

		N
NBCT01	0	16093
	1	172

a. Grade = 3

Descriptive Statistics[a]

Dependent Variable: LGN01

NBCT01	Mean	Std. Deviation	N
0	28.8977	28.17418	16093
1	30.4244	29.09312	172
Total	28.9138	28.18357	16265

a. Grade = 3

Tests of Between-Subjects Effects[b]

Dependent Variable: LGN01

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	465653.108[a]	2	232826.554	304.040	.000
Intercept	753643.476	1	753643.476	984.156	.000
LANGSS00	465256.416	1	465256.416	607.561	.000
NBCT01	1344.879	1	1344.879	1.756	.185
Error	12453062.0	16262	765.777		
Total	26516385.0	16265			
Corrected Total	12918715.2	16264			

a. R Squared = .036 (Adjusted R Squared = .036)

b. Grade = 3

Estimated Marginal Means

NBCT01[b]

Dependent Variable: LGN01

NBCT01	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	28.884[a]	.218	28.456	29.312
1	31.696[a]	2.111	27.559	35.833

a. Covariates appearing in the model are evaluated at the following values: L scale score = 573.15.

b. Grade = 3

Grade = 4

Between-Subjects Factors[a]

		N
NBCT01	0	16742
	1	154

a. Grade = 4

Descriptive Statistics[a]

Dependent Variable: LGN01

NBCT01	Mean	Std. Deviation	N
0	20.0923	25.42450	16742
1	19.3571	22.81259	154
Total	20.0856	25.40139	16896

a. Grade = 4

Tests of Between-Subjects Effects[b]

Dependent Variable: LGN01

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3364860.258[a]	2	1682430.129	3771.247	.000
Intercept	3582525.652	1	3582525.652	8030.402	.000
LANGSS00	3364777.777	1	3364777.777	7542.310	.000
NBCT01	7.325	1	7.325	.016	.898
Error	7536310.819	16893	446.120		
Total	17717575.0	16896			
Corrected Total	10901171.1	16895			

a. R Squared = .309 (Adjusted R Squared = .309)

b. Grade = 4

Estimated Marginal Means

NBCT01[b]

Dependent Variable: LGN01

NBCT01	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	20.084[a]	.163	19.764	20.404
1	20.303[a]	1.702	16.967	23.639

a. Covariates appearing in the model are evaluated at the following values: L scale score = 600.65.

b. Grade = 4

Grade = 5

Between-Subjects Factors[a]

		N
NBCT01	0	17159
	1	87

a. Grade = 5

Descriptive Statistics[a]

Dependent Variable: LGN01

NBCT01	Mean	Std. Deviation	N
0	11.2985	22.40306	17159
1	13.3448	21.77412	87
Total	11.3088	22.39979	17246

a. Grade = 5

Tests of Between-Subjects Effects[b]

Dependent Variable: LGN01

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	932242.543[a]	2	466121.272	1041.045	.000
Intercept	968450.929	1	968450.929	2162.958	.000
LANGSS00	931880.073	1	931880.073	2081.280	.000
NBCT01	1355.897	1	1355.897	3.028	.082
Error	7720446.654	17243	447.744		
Total	10858272.0	17246			
Corrected Total	8652689.197	17245			

a. R Squared = .108 (Adjusted R Squared = .108)

b. Grade = 5

Estimated Marginal Means

NBCT01[b]

Dependent Variable: LGN01

NBCT01	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	11.289[a]	.162	10.972	11.605
1	15.247[a]	2.269	10.800	19.695

a. Covariates appearing in the model are evaluated at the following values: L scale score = 619.09.

b. Grade = 5

Grade = 6

Between-Subjects Factors[a]

		N
NBCT01	0	12773
	1	65

a. Grade = 6

Descriptive Statistics[a]

Dependent Variable: LGN01

NBCT01	Mean	Std. Deviation	N
0	12.9796	21.96361	12773
1	15.0769	18.72274	65
Total	12.9903	21.94829	12838

a. Grade = 6

Tests of Between-Subjects Effects[b]

Dependent Variable: LGN01

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	958738.248[a]	2	479369.124	1177.507	.000
Intercept	991696.387	1	991696.387	2435.971	.000
LANGSS00	958453.788	1	958453.788	2354.315	.000
NBCT01	3310.277	1	3310.277	8.131	.004
Error	5225195.535	12835	407.105		
Total	8350307.000	12838			
Corrected Total	6183933.783	12837			

a. R Squared = .155 (Adjusted R Squared = .155)

b. Grade = 6

Estimated Marginal Means

NBCT01[b]

Dependent Variable: LGN01

NBCT01	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	12.954[a]	.179	12.604	13.304
1	20.115[a]	2.505	15.205	25.024

a. Covariates appearing in the model are evaluated at the following values: L scale score = 628.69.

b. Grade = 6

2001-2002 - READING

Grade = 3

Between-Subjects Factors[a]

		N
NBCT02	0	15339
	1	144

a. Grade = 3

Descriptive Statistics[a]

Dependent Variable: RGN02

NBCT02	Mean	Std. Deviation	N
0	32.1610	26.53326	15339
1	32.8958	25.79064	144
Total	32.1678	26.52573	15483

a. Grade = 3

Tests of Between-Subjects Effects[b]

Dependent Variable: RGN02

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	658993.615[a]	2	329496.807	498.381	.000
Intercept	1040092.993	1	1040092.993	1573.194	.000
READSS01	658916.573	1	658916.573	996.645	.000
NBCT02	39.783	1	39.783	.060	.806
Error	10234366.4	15480	661.135		
Total	26914660.0	15483			
Corrected Total	10893360.1	15482			

a. R Squared = .060 (Adjusted R Squared = .060)

b. Grade = 3

Estimated Marginal Means

NBCT02[b]

Dependent Variable: RGN02

NBCT02	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	32.163[a]	.208	31.756	32.570
1	32.691[a]	2.143	28.491	36.891

a. Covariates appearing in the model are evaluated at the following values: R scale score = 600.31.

b. Grade = 3

Grade = 4

Between-Subjects Factors[a]

		N
NBCT02	0	16178
	1	114

a. Grade = 4

Descriptive Statistics[a]

Dependent Variable: RGN02

NBCT02	Mean	Std. Deviation	N
0	28.1797	23.94624	16178
1	31.3596	23.81943	114
Total	28.2019	23.94609	16292

a. Grade = 4

Tests of Between-Subjects Effects[b]

Dependent Variable: RGN02

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1194953.051[a]	2	597476.526	1194.651	.000
Intercept	1560950.513	1	1560950.513	3121.113	.000
READSS01	1193808.332	1	1193808.332	2387.014	.000
NBCT02	943.968	1	943.968	1.887	.170
Error	8146556.568	16289	500.126		
Total	22299342.0	16292			
Corrected Total	9341509.619	16291			

a. R Squared = .128 (Adjusted R Squared = .128)

b. Grade = 4

Estimated Marginal Means

NBCT02[b]

Dependent Variable: RGN02

NBCT02	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	28.182[a]	.176	27.837	28.526
1	31.069[a]	2.095	26.964	35.175

a. Covariates appearing in the model are evaluated at the following values: R scale score = 629.92.

b. Grade = 4

Grade = 5

Between-Subjects Factors[a]

		N
NBCT02	0	17029
	1	60

a. Grade = 5

Descriptive Statistics[a]

Dependent Variable: RGN02

NBCT02	Mean	Std. Deviation	N
0	12.1990	22.25402	17029
1	6.7500	22.54684	60
Total	12.1799	22.25672	17089

a. Grade = 5

Tests of Between-Subjects Effects[b]

Dependent Variable: RGN02

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1972425.142[a]	2	986212.571	2595.442	.000
Intercept	1729132.206	1	1729132.206	4550.603	.000
READSS01	1970649.892	1	1970649.892	5186.212	.000
NBCT02	449.802	1	449.802	1.184	.277
Error	6492315.902	17086	379.979		
Total	10999886.0	17089			
Corrected Total	8464741.043	17088			

a. R Squared = .233 (Adjusted R Squared = .233)

b. Grade = 5

NBCT02[b]

Dependent Variable: RGN02

NBCT02	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	12.190[a]	.149	11.897	12.482
1	9.446[a]	2.517	4.513	14.380

a. Covariates appearing in the model are evaluated at the following values: R scale score = 657.49.

b. Grade = 5

Estimated Marginal Means

Grade = 6

Between-Subjects Factors[a]

		N
NBCT02	0	14421
	1	81

a. Grade = 6

Descriptive Statistics[a]

Dependent Variable: RGN02

NBCT02	Mean	Std. Deviation	N
0	12.3209	19.95089	14421
1	7.8148	15.47749	81
Total	12.2957	19.93110	14502

a. Grade = 6

Tests of Between-Subjects Effects[b]

Dependent Variable: RGN02

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1051974.102[a]	2	525987.051	1619.674	.000
Intercept	1054564.768	1	1054564.768	3247.325	.000
READSS01	1050338.635	1	1050338.635	3234.312	.000
NBCT02	90.613	1	90.613	.279	.597
Error	4708532.007	14499	324.749		
Total	7952974.000	14502			
Corrected Total	5760506.110	14501			

a. R Squared = .183 (Adjusted R Squared = .183)

b. Grade = 6

Estimated Marginal Means

NBCT02[b]

Dependent Variable: RGN02

NBCT02	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	12.302[a]	.150	12.007	12.596
1	11.240[a]	2.003	7.314	15.167

a. Covariates appearing in the model are evaluated at the following values: R scale score = 667.80.

b. Grade = 6

2001-2002 - MATH

Grade = 3

Between-Subjects Factors[a]

		N
NBCT02	0	15973
	1	150

a. Grade = 3

Descriptive Statistics[a]

Dependent Variable: MGN02

NBCT02	Mean	Std. Deviation	N
0	26.3672	28.93695	15973
1	26.0533	28.82015	150
Total	26.3643	28.93499	16123

a. Grade = 3

Tests of Between-Subjects Effects[b]

Dependent Variable: MGN02

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1516881.224[a]	2	758440.612	1020.454	.000
Intercept	1886727.785	1	1886727.785	2538.523	.000
MATHSS01	1516866.580	1	1516866.580	2040.889	.000
NBCT02	22.364	1	22.364	.030	.862
Error	11981000.7	16120	743.238		
Total	24704618.0	16123			
Corrected Total	13497882.0	16122			

a. R Squared = .112 (Adjusted R Squared = .112)

b. Grade = 3

Estimated Marginal Means

NBCT02[b]

Dependent Variable: MGN02

NBCT02	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	26.361[a]	.216	25.938	26.784
1	26.749[a]	2.226	22.385	31.112

a. Covariates appearing in the model are evaluated at the following values: M scale score = 591.33.

b. Grade = 3

Grade = 4

Between-Subjects Factors[a]

		N
NBCT02	0	16640
	1	116

a. Grade = 4

Descriptive Statistics[a]

Dependent Variable: MGN02

NBCT02	Mean	Std. Deviation	N
0	30.9508	25.88085	16640
1	30.1810	26.68483	116
Total	30.9455	25.88576	16756

a. Grade = 4

Tests of Between-Subjects Effects[b]

Dependent Variable: MGN02

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1768272.476[a]	2	884136.238	1565.943	.000
Intercept	2185335.298	1	2185335.298	3870.570	.000
MATHSS01	1768204.220	1	1768204.220	3131.766	.000
NBCT02	4.023	1	4.023	.007	.933
Error	9458793.668	16753	564.603		
Total	27272964.0	16756			
Corrected Total	11227066.1	16755			

a. R Squared = .158 (Adjusted R Squared = .157)

b. Grade = 4

Estimated Marginal Means

NBCT02[b]

Dependent Variable: MGN02

NBCT02	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	30.947[a]	.184	30.586	31.308
1	30.760[a]	2.206	26.435	35.084

a. Covariates appearing in the model are evaluated at the following values: M scale score = 615.44.

b. Grade = 4

Grade = 5

Between-Subjects Factors[a]

		N
NBCT02	0	17506
	1	61

a. Grade = 5

Descriptive Statistics[a]

Dependent Variable: MGN02

NBCT02	Mean	Std. Deviation	N
0	25.3002	24.13303	17506
1	25.6066	23.86509	61
Total	25.3012	24.13144	17567

a. Grade = 5

Tests of Between-Subjects Effects[b]

Dependent Variable: MGN02

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	766117.140[a]	2	383058.570	710.982	.000
Intercept	922463.383	1	922463.383	1712.152	.000
MATHSS01	766111.434	1	766111.434	1421.953	.000
NBCT02	435.439	1	435.439	.808	.369
Error	9463030.663	17564	538.774		
Total	21474717.0	17567			
Corrected Total	10229147.8	17566			

a. R Squared = .075 (Adjusted R Squared = .075)

b. Grade = 5

Estimated Marginal Means

NBCT02[b]

Dependent Variable: MGN02

NBCT02	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	25.292[a]	.175	24.948	25.636
1	27.969[a]	2.973	22.142	33.796

a. Covariates appearing in the model are evaluated at the following values: M scale score = 644.82.

b. Grade = 5

Grade = 6

Between-Subjects Factors[a]

		N
NBCT02	0	14821
	1	80

a. Grade = 6

Descriptive Statistics[a]

Dependent Variable: MGN02

NBCT02	Mean	Std. Deviation	N
0	20.7508	22.89775	14821
1	22.6125	20.20558	80
Total	20.7608	22.88394	14901

a. Grade = 6

Tests of Between-Subjects Effects[b]

Dependent Variable: MGN02

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	478799.403[a]	2	239399.702	486.974	.000
Intercept	615610.648	1	615610.648	1252.242	.000
MATHSS01	478523.606	1	478523.606	973.387	.000
NBCT02	1488.495	1	1488.495	3.028	.082
Error	7323956.686	14898	491.607		
Total	14225220.0	14901			
Corrected Total	7802756.089	14900			

a. R Squared = .061 (Adjusted R Squared = .061)

b. Grade = 6

Estimated Marginal Means

NBCT02[b]

Dependent Variable: MGN02

NBCT02	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	20.738[a]	.182	20.381	21.095
1	25.065[a]	2.480	20.203	29.926

a. Covariates appearing in the model are evaluated at the following values: M scale score = 667.26.

b. Grade = 6

2001-2002 - LANGUAGE

Grade = 3

Between-Subjects Factors[a]

		N
NBCT02	0	15951
	1	153

a. Grade = 3

Descriptive Statistics[a]

Dependent Variable: LGN02

NBCT02	Mean	Std. Deviation	N
0	30.0347	28.30334	15951
1	27.0980	30.59084	153
Total	30.0068	28.32635	16104

a. Grade = 3

Tests of Between-Subjects Effects[b]

Dependent Variable: LGN02

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	551460.214[a]	2	275730.107	358.915	.000
Intercept	838346.651	1	838346.651	1091.268	.000
LANGSS01	550153.253	1	550153.253	716.129	.000
NBCT02	1279.092	1	1279.092	1.665	.197
Error	12369297.0	16101	768.232		
Total	27420958.0	16104			
Corrected Total	12920757.2	16103			

a. R Squared = .043 (Adjusted R Squared = .043)

b. Grade = 3

Estimated Marginal Means

NBCT02[b]

Dependent Variable: LGN02

NBCT02	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	30.034[a]	.219	29.604	30.465
1	27.129[a]	2.241	22.737	31.521

a. Covariates appearing in the model are evaluated at the following values: L scale score = 573.78.

b. Grade = 3

Grade = 4

Between-Subjects Factors[a]

		N
NBCT02	0	16552
	1	113

a. Grade = 4

Descriptive Statistics[a]

Dependent Variable: LGN02

NBCT02	Mean	Std. Deviation	N
0	19.9959	25.37173	16552
1	15.8053	21.19588	113
Total	19.9675	25.34753	16665

a. Grade = 4

Tests of Between-Subjects Effects[b]

Dependent Variable: LGN02

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3268724.764[a]	2	1634362.382	3661.238	.000
Intercept	3309037.715	1	3309037.715	7412.784	.000
LANGSS01	3266753.830	1	3266753.830	7318.061	.000
NBCT02	547.657	1	547.657	1.227	.268
Error	7437851.608	16662	446.396		
Total	17350914.0	16665			
Corrected Total	10706576.4	16664			

a. R Squared = .305 (Adjusted R Squared = .305)

b. Grade = 4

Estimated Marginal Means

NBCT02[b]

Dependent Variable: LGN02

NBCT02	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	19.982[a]	.164	19.661	20.304
1	17.773[a]	1.988	13.877	21.669

a. Covariates appearing in the model are evaluated at the following values: L scale score = 601.97.

b. Grade = 4

Grade = 5

Between-Subjects Factors[a]

		N
NBCT02	0	17384
	1	62

a. Grade = 5

Descriptive Statistics[a]

Dependent Variable: LGN02

NBCT02	Mean	Std. Deviation	N
0	11.5579	22.58502	17384
1	11.2581	24.35996	62
Total	11.5568	22.59083	17446

a. Grade = 5

Tests of Between-Subjects Effects[b]

Dependent Variable: LGN02

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	928551.204[a]	2	464275.602	1015.541	.000
Intercept	908048.484	1	908048.484	1986.235	.000
LANGSS01	928545.651	1	928545.651	2031.070	.000
NBCT02	113.565	1	113.565	.248	.618
Error	7974428.004	17443	457.171		
Total	11233062.0	17446			
Corrected Total	8902979.207	17445			

a. R Squared = .104 (Adjusted R Squared = .104)

b. Grade = 5

Estimated Marginal Means

NBCT02[b]

Dependent Variable: LGN02

NBCT02	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	11.552[a]	.162	11.234	11.870
1	12.908[a]	2.716	7.585	18.231

a. Covariates appearing in the model are evaluated at the following values: L scale score = 620.62.

b. Grade = 5

Grade = 6

Between-Subjects Factors[a]

		N
NBCT02	0	14651
	1	78

a. Grade = 6

Descriptive Statistics[a]

Dependent Variable: LGN02

NBCT02	Mean	Std. Deviation	N
0	13.8307	21.92071	14651
1	10.8205	25.78570	78
Total	13.8147	21.94303	14729

a. Grade = 6

Tests of Between-Subjects Effects[b]

Dependent Variable: LGN02

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	979725.499[a]	2	489862.749	1180.302	.000
Intercept	1009679.375	1	1009679.375	2432.776	.000
LANGSS01	979022.485	1	979022.485	2358.910	.000
NBCT02	135.880	1	135.880	.327	.567
Error	6111757.870	14726	415.032		
Total	9902461.000	14729			
Corrected Total	7091483.369	14728			

a. R Squared = .138 (Adjusted R Squared = .138)

b. Grade = 6

Estimated Marginal Means

NBCT02[b]

Dependent Variable: LGN02

NBCT02	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	13.808[a]	.168	13.478	14.138
1	15.132[a]	2.308	10.607	19.657

a. Covariates appearing in the model are evaluated at the following values: L scale score = 629.73.

b. Grade = 6

2002-2003 - READING

Grade = 3

Between-Subjects Factors[a]

		N
NBCT03	0	15541
	1	105

a. GRADE03 = 3

Descriptive Statistics[a]

Dependent Variable: RGN03

NBCT03	Mean	Std. Deviation	N
0	31.1703	26.18790	15541
1	37.2476	25.30992	105
Total	31.2111	26.18603	15646

a. GRADE03 = 3

Tests of Between-Subjects Effects[b]

Dependent Variable: RGN03

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	718655.351[a]	2	359327.675	561.577	.000
Intercept	1102789.546	1	1102789.546	1723.500	.000
READSS02	714803.357	1	714803.357	1117.134	.000
NBCT03	5694.888	1	5694.888	8.900	.003
Error	10009246.4	15643	639.855		
Total	25969191.0	15646			
Corrected Total	10727901.7	15645			

a. R Squared = .067 (Adjusted R Squared = .067)

b. GRADE03 = 3

Estimated Marginal Means

NBCT03[b]

Dependent Variable: RGN03

NBCT03	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	31.162[a]	.203	30.764	31.559
1	38.552[a]	2.469	33.713	43.391

a. Covariates appearing in the model are evaluated at the following values: R scale score = 598.92.

b. GRADE03 = 3

Grade = 4

Between-Subjects Factors[a]

		N
NBCT03	0	16103
	1	123

a. GRADE03 = 4

Descriptive Statistics[a]

Dependent Variable: RGN03

NBCT03	Mean	Std. Deviation	N
0	27.9769	24.13572	16103
1	31.6585	22.43089	123
Total	28.0048	24.12472	16226

a. GRADE03 = 4

Tests of Between-Subjects Effects[b]

Dependent Variable: RGN03

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1415359.625[a]	2	707679.813	1430.148	.000
Intercept	1818589.319	1	1818589.319	3675.182	.000
READSS02	1413705.065	1	1413705.065	2856.953	.000
NBCT03	2130.865	1	2130.865	4.306	.038
Error	8027622.000	16223	494.830		
Total	22168534.0	16226			
Corrected Total	9442981.625	16225			

a. R Squared = .150 (Adjusted R Squared = .150)

b. GRADE03 = 4

Estimated Marginal Means

NBCT03[b]

Dependent Variable: RGN03

NBCT03	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	27.973[a]	.175	27.630	28.317
1	32.151[a]	2.006	28.220	36.083

a. Covariates appearing in the model are evaluated at the following values: R scale score = 628.45.

b. GRADE03 = 4

Grade = 5

Between-Subjects Factors[a]

		N
NBCT03	0	16565
	1	40

a. GRADE03 = 5

Descriptive Statistics[a]

Dependent Variable: RGN03

NBCT03	Mean	Std. Deviation	N
0	12.7717	22.46121	16565
1	11.9750	25.02459	40
Total	12.7698	22.46693	16605

a. GRADE03 = 5

Tests of Between-Subjects Effects[b]

Dependent Variable: RGN03

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1961826.012[a]	2	980913.006	2536.916	.000
Intercept	1599474.668	1	1599474.668	4136.690	.000
READSS02	1961800.685	1	1961800.685	5073.767	.000
NBCT03	104.378	1	104.378	.270	.603
Error	6419256.812	16602	386.656		
Total	11088810.0	16605			
Corrected Total	8381082.824	16604			

a. R Squared = .234 (Adjusted R Squared = .234)

b. GRADE03 = 5

Estimated Marginal Means

NBCT03[b]

Dependent Variable: RGN03

NBCT03	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	12.766[a]	.153	12.466	13.065
1	14.383[a]	3.109	8.289	20.478

a. Covariates appearing in the model are evaluated at the following values: R scale score = 654.93.

b. GRADE03 = 5

Grade = 6

Between-Subjects Factors[a]

		N
NBCT03	0	12743
	1	84

a. GRADE03 = 6

Descriptive Statistics[a]

Dependent Variable: RGN03

NBCT03	Mean	Std. Deviation	N
0	12.1601	19.78339	12743
1	10.6310	15.69881	84
Total	12.1501	19.75929	12827

a. GRADE03 = 6

Tests of Between-Subjects Effects[b]

Dependent Variable: RGN03

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	891656.624[a]	2	445828.312	1389.046	.000
Intercept	924430.887	1	924430.887	2880.206	.000
READSS02	891461.496	1	891461.496	2777.484	.000
NBCT03	281.963	1	281.963	.878	.349
Error	4115991.484	12824	320.960		
Total	6901225.000	12827			
Corrected Total	5007648.107	12826			

a. R Squared = .178 (Adjusted R Squared = .178)

b. GRADE03 = 6

Estimated Marginal Means

NBCT03[b]

Dependent Variable: RGN03

NBCT03	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	12.138[a]	.159	11.827	12.449
1	13.977[a]	1.956	10.144	17.811

a. Covariates appearing in the model are evaluated at the following values: R scale score = 665.59.

b. GRADE03 = 6

2002-2003 – MATH

Grade = 3

Between-Subjects Factors[a]

		N
NBCT03	0	16275
	1	118

a. GRADE03 = 3

Descriptive Statistics[a]

Dependent Variable: MGN03

NBCT03	Mean	Std. Deviation	N
0	27.3004	29.37467	16275
1	33.3898	27.87182	118
Total	27.3442	29.36783	16393

a. GRADE03 = 3

Tests of Between-Subjects Effects[b]

Dependent Variable: MGN03

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1570820.371[a]	2	785410.186	1024.357	.000
Intercept	1979936.756	1	1979936.756	2582.297	.000
MATHSS02	1566476.289	1	1566476.289	2043.049	.000
NBCT03	4795.116	1	4795.116	6.254	.012
Error	12566780.1	16390	766.735		
Total	26394762.0	16393			
Corrected Total	14137600.5	16392			

a. R Squared = .111 (Adjusted R Squared = .111)

b. GRADE03 = 3

Estimated Marginal Means

NBCT03[b]

Dependent Variable: MGN03

NBCT03	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	27.298[a]	.217	26.873	27.724
1	33.696[a]	2.549	28.699	38.692

a. Covariates appearing in the model are evaluated at the following values: M scale score = 589.59.

b. GRADE03 = 3

Grade = 4

Between-Subjects Factors[a]

		N
NBCT03	0	16582
	1	127

a. GRADE03 = 4

Descriptive Statistics[a]

Dependent Variable: MGN03

NBCT03	Mean	Std. Deviation	N
0	31.3779	26.04577	16582
1	38.3150	26.81290	127
Total	31.4307	26.05783	16709

a. GRADE03 = 4

Tests of Between-Subjects Effects[b]

Dependent Variable: MGN03

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1891112.733[a]	2	945556.367	1670.912	.000
Intercept	2443563.124	1	2443563.124	4318.071	.000
MATHSS02	1885047.658	1	1885047.658	3331.107	.000
NBCT03	7599.187	1	7599.187	13.429	.000
Error	9453796.194	16706	565.892		
Total	27851509.0	16709			
Corrected Total	11344908.9	16708			

a. R Squared = .167 (Adjusted R Squared = .167)

b. GRADE03 = 4

Estimated Marginal Means

NBCT03[b]

Dependent Variable: MGN03

NBCT03	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	31.372[a]	.185	31.010	31.734
1	39.137[a]	2.111	34.999	43.274

a. Covariates appearing in the model are evaluated at the following values: M scale score = 614.65.

b. GRADE03 = 4

Grade = 5

Between-Subjects Factors[a]

		N
NBCT03	0	17020
	1	41

a. GRADE03 = 5

Descriptive Statistics[a]

Dependent Variable: MGN03

NBCT03	Mean	Std. Deviation	N
0	25.2879	24.33578	17020
1	21.7073	29.74999	41
Total	25.2793	24.34980	17061

a. GRADE03 = 5

Tests of Between-Subjects Effects[b]

Dependent Variable: MGN03

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	991406.227[a]	2	495703.113	926.786	.000
Intercept	1010126.947	1	1010126.947	1888.573	.000
MATHSS02	990881.847	1	990881.847	1852.592	.000
NBCT03	60.778	1	60.778	.114	.736
Error	9123683.947	17058	534.862		
Total	21017796.0	17061			
Corrected Total	10115090.2	17060			

a. R Squared = .098 (Adjusted R Squared = .098)

b. GRADE03 = 5

Estimated Marginal Means

NBCT03[b]

Dependent Variable: MGN03

NBCT03	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	25.282[a]	.177	24.935	25.630
1	24.063[a]	3.612	16.983	31.143

a. Covariates appearing in the model are evaluated at the following values: M scale score = 643.82.

b. GRADE03 = 5

Grade = 6

Between-Subjects Factors[a]

		N
NBCT03	0	12984
	1	84

a. GRADE03 = 6

Descriptive Statistics[a]

Dependent Variable: MGN03

NBCT03	Mean	Std. Deviation	N
0	19.8696	22.37597	12984
1	20.2262	26.87739	84
Total	19.8719	22.40658	13068

a. GRADE03 = 6

Tests of Between-Subjects Effects[b]

Dependent Variable: MGN03

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	420132.390[a]	2	210066.195	446.973	.000
Intercept	538474.212	1	538474.212	1145.752	.000
MATHSS02	420121.778	1	420121.778	893.924	.000
NBCT03	820.648	1	820.648	1.746	.186
Error	6140219.172	13065	469.975		
Total	11720806.0	13068			
Corrected Total	6560351.562	13067			

a. R Squared = .064 (Adjusted R Squared = .064)

b. GRADE03 = 6

Estimated Marginal Means

NBCT03[b]

Dependent Variable: MGN03

NBCT03	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	19.852[a]	.190	19.479	20.225
1	22.990[a]	2.367	18.350	27.630

a. Covariates appearing in the model are evaluated at the following values: M scale score = 666.31.

b. GRADE03 = 6

2002-2003 – LANGUAGE

Grade = 3

Between-Subjects Factors[a]

		N
NBCT03	0	16144
	1	117

a. GRADE03 = 3

Descriptive Statistics[a]

Dependent Variable: LGN03

NBCT03	Mean	Std. Deviation	N
0	29.9963	28.43310	16144
1	33.3846	28.87752	117
Total	30.0207	28.43686	16261

a. GRADE03 = 3

Tests of Between-Subjects Effects[b]

Dependent Variable: LGN03

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	687501.305[a]	2	343750.653	448.487	.000
Intercept	1022849.478	1	1022849.478	1334.498	.000
LANGSS02	686167.766	1	686167.766	895.234	.000
NBCT03	2471.842	1	2471.842	3.225	.073
Error	12461232.7	16258	766.468		
Total	27803861.0	16261			
Corrected Total	13148734.0	16260			

a. R Squared = .052 (Adjusted R Squared = .052)

b. GRADE03 = 3

Estimated Marginal Means

NBCT03[b]

Dependent Variable: LGN03

NBCT03	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	29.988[a]	.218	29.560	30.415
1	34.601[a]	2.560	29.584	39.619

a. Covariates appearing in the model are evaluated at the following values: L scale score = 572.49.

b. GRADE03 = 3

Grade = 4

Between-Subjects Factors[a]

		N
NBCT03	0	16529
	1	128

a. GRADE03 = 4

Descriptive Statistics[a]

Dependent Variable: LGN03

NBCT03	Mean	Std. Deviation	N
0	19.5912	25.84293	16529
1	24.4844	30.43878	128
Total	19.6288	25.88381	16657

a. GRADE03 = 4

Tests of Between-Subjects Effects[b]

Dependent Variable: LGN03

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3673065.345[a]	2	1836532.673	4085.718	.000
Intercept	3862039.978	1	3862039.978	8591.847	.000
LANGSS02	3670024.176	1	3670024.176	8164.671	.000
NBCT03	4033.013	1	4033.013	8.972	.003
Error	7485982.554	16654	449.501		
Total	17576823.0	16657			
Corrected Total	11159047.9	16656			

a. R Squared = .329 (Adjusted R Squared = .329)

b. GRADE03 = 4

Estimated Marginal Means

NBCT03[b]

Dependent Variable: LGN03

NBCT03	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	19.586[a]	.165	19.262	19.909
1	25.220[a]	1.874	21.547	28.894

a. Covariates appearing in the model are evaluated at the following values: L scale score = 600.70.

b. GRADE03 = 4

Grade = 5

Between-Subjects Factors[a]

		N
NBCT03	0	16905
	1	43

a. GRADE03 = 5

Descriptive Statistics[a]

Dependent Variable: LGN03

NBCT03	Mean	Std. Deviation	N
0	11.3230	22.54932	16905
1	4.9302	20.92435	43
Total	11.3068	22.54707	16948

a. GRADE03 = 5

Tests of Between-Subjects Effects[b]

Dependent Variable: LGN03

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1023158.914[a]	2	511579.457	1141.793	.000
Intercept	881850.901	1	881850.901	1968.202	.000
LANGSS02	1021406.049	1	1021406.049	2279.674	.000
NBCT03	878.627	1	878.627	1.961	.161
Error	7592191.618	16945	448.049		
Total	10782054.0	16948			
Corrected Total	8615350.532	16947			

a. R Squared = .119 (Adjusted R Squared = .119)

b. GRADE03 = 5

Estimated Marginal Means

NBCT03[b]

Dependent Variable: LGN03

NBCT03	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	11.318[a]	.163	10.999	11.637
1	6.792[a]	3.228	.464	13.120

a. Covariates appearing in the model are evaluated at the following values: L scale score = 619.67.

b. GRADE03 = 5

Grade = 6

Between-Subjects Factors[a]

		N
NBCT03	0	12914
	1	84

a. GRADE03 = 6

Descriptive Statistics[a]

Dependent Variable: LGN03

NBCT03	Mean	Std. Deviation	N
0	13.3690	22.06348	12914
1	13.1667	22.03329	84
Total	13.3677	22.06244	12998

a. GRADE03 = 6

Tests of Between-Subjects Effects[b]

Dependent Variable: LGN03

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	823396.407[a]	2	411698.204	972.216	.000
Intercept	879010.885	1	879010.885	2075.765	.000
LANGSS02	823392.991	1	823392.991	1944.424	.000
NBCT03	881.236	1	881.236	2.081	.149
Error	5502909.489	12995	423.464		
Total	8648979.000	12998			
Corrected Total	6326305.896	12997			

a. R Squared = .130 (Adjusted R Squared = .130)

b. GRADE03 = 6

Estimated Marginal Means

NBCT03[b]

Dependent Variable: LGN03

NBCT03	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	13.347[a]	.181	12.992	13.702
1	16.598[a]	2.247	12.194	21.002

a. Covariates appearing in the model are evaluated at the following values: L scale score = 628.56.

b. GRADE03 = 6