Value-Added Assessment of Teacher Quality As an Alternative to the National Board for Professional Teaching Standards: What Recent Studies Say

George K. Cunningham, PhD
University of Louisville

J. E. Stone, EdD
East Tennessee State University

Abstract

Virtually every state is committed to increasing student achievement as measured by improved standardized test scores. Most of them are also encouraging teachers to become certified by the National Board for Professional Teaching Standards (NBPTS)--an organization that embraces a different set of educational priorities.

Four recent studies have examined the annual increases in student achievement produced by NBPTS teachers and found them to be slightly larger than the average of their non-certified peers. However, relative to the achievement gaps identified by the No Child Left Behind Act, the advantage associated with NBPTS certification is trivial.

Although not designed as reports on the value-added performance of non-certified teachers, the recent studies show that the top 10 percent of non-certified teachers produce achievement effect sizes 10 to 20 times greater than those produced by the average NBPTS-certified teacher.

Gains of this magnitude are substantial relative to the learning gaps revealed by NCLB. Teachers producing such gains could be identified by the kind of value-added analysis used in the present studies. These findings suggest that the goal of improved student achievement would be better served by a program of bonuses for student achievement gains than by one of rewards for NBPTS certification.

Cite as:

Cunningham, G. K. & Stone, J. E. (2005). Value-added assessment of teacher quality as an alternative to the National Board for Professional Teaching Standards: What recent studies say. In Robert Lissitz (Ed.). *Value added models in education: Theory and applications*. Maple Grove, MN: JAM Press.

Governors and state legislatures in every state have embarked on education reform programs that are strongly supported by the public. All of these programs emphasize the goal of improved academic achievement operationally defined as student performance on standardized achievement tests.

In contrast to this trend, some of the nation's most important education organizations have promoted teaching standards that emphasize a different set of educational priorities. They include the National Board for Professional Teaching Standards (NBPTS) and related groups such as the National Commission on Teaching and America's Future (NCTAF), the National Council for the Accreditation of Teacher Education (NCATE), and the Interstate New Teacher Assessment and Support Consortium (INTASC).

The result has been an ongoing clash between two educational cultures: (1) those who believe that the most important activity of schools is the enhancement of objectively measured academic achievement, and (2) those who believe test scores are narrow and artificial indicators of learning and therefore should not be treated as education's prime objective.

According to the latter culture, true learning is evidenced only by real performance in the real world. In fact, to those who embrace the second culture, reform focused on the improvement of test scores is undesirable because it encourages the use of result-oriented teaching methods, not the process-oriented approaches that they believe are better suited to critical thinking (Casas, 2003; Smerdon, Burkam, & Lee, 1999). Unfortunately, the portfolios and authentic assessments that are suited to real-world (i.e., not merely classroom) performance are far more subjective and less reliable than the standardized tests that they would replace. In effect, the second culture would rather water down accountability than temper its pedagogical idealism.

Schools of education operating under the auspices of NCATE and INTASC have embraced the second of the two cultures. They train teachers as though their states have not placed a high premium on school and teacher accountability for test performance. The same holds true for the NBPTS certification program that has now been adopted in thirty states. It places little importance on academic achievement as measured by standardized achievement tests.

NBPTS's Five Propositions of Accomplished Teaching say that teachers should be aware of the "broad goals, objectives, and priorities" set by authorities and of their legal obligation to carry out public policy. However, they also suggest that teachers should consult with colleagues and make their own decisions about what students should learn (National Board for Professional Teaching Standards, 2005, p.18). In truth, the Five Propositions give little attention to teacher accountability for student achievement, and even that limited discussion is diluted by a variety of caveats.

A more forthright expression of the NBPTS viewpoint is revealed in the first major validation study commissioned by NBPTS (Bond, Smith, Baker, & Hattie, 2000):

> Brief additional mention should also be made of the deliberate design decision in the present investigation to use measures of student achievement other than commercially or state-developed multiple-choice tests of generic academic subjects such as reading and mathematics. It is not too much of an exaggeration to state that such measures have been cited as the cause of all of the nation's considerable problems in educating our youth. To be sure, the overuse and misuse of multiple-choice tests is well documented. (p. 141)

This is an astounding statement. The authors of this study are not merely saying that there are problems with the use of standardized tests and that they prefer to use other methods, they are suggesting that the use of commercial and state-developed multiple-choice tests is the cause of the nation's educational problems. To NBPTS, a superior teacher is one committed to a student-centered, constructivist style of instruction, regardless of whether the use of such practices produces gains in objectively measured student achievement (Ballou, 2003).

Most states seem unaware of the discrepancy and are committed to expanding NBPTS participation. They are investing millions on a program that they hope will improve academic achievement when, in fact, the program philosophy disagrees with the aim of academic achievement as an unrivaled educational priority.

Over the past five years, several studies have provided evidence that sheds light on the question of how well the aim of improved achievement is served by NBPTS certification. The present report examines their implications for public policy by (1) evaluating and synthesizing their findings and (2) by comparing those findings to the achievement gains obtainable through value-added assessment of teacher performance.

NBPTS Validity: Early Studies

During the 1990s, the Department of Education provided millions of dollars for NBPTS with the understanding its effectiveness would eventually be established. In 1996, Buday and Kelly described the National Board's early efforts to measure the impact of assessment on candidates for certification: "Preliminary data are in the form of anecdotes and testimonials from candidates, virtually all of whom report that the process offers tremendous potential for improving student learning (p. 217)." While this finding demonstrates that it is possible to identify candidates who speak well of the assessment process, it is not evidence that NBPTS certified teachers are exceptionally effective in bringing about student achievement.

The best known attempt to validate NBPTS certification (Bond, Smith, Baker, & Hattie, 2000) similarly failed to examine linkages to achievement test scores. It compared 31 certified teachers with 34 others who unsuccessfully applied for Board certification. Classroom observations were used to determine the degree to which the certified and uncertified teachers adhered to NBPTS principles. Student achievement was estimated by each teacher reviewing a portfolio of student work products and writing samples.

The most noteworthy finding of the study was that the teacher groups differed on 11 of the 13 classroom behaviors that were said to reflect the NBPTS teaching propositions. Given that the NBPTS teachers were certified on the basis of their adherence to the NBPTS propositions, the finding must be seen as predictable.

The students taught by the Board-certified teachers scored slightly higher on the quality of their portfolios, but their superiority may have been due to preexisting differences in achievement (Podgursky, 2001). There were no differences between the two groups with regard to writing performance.

In reality, what Bond, Smith, Baker, & Hattie showed is that Board Certified teachers persist in exhibiting the behaviors and beliefs that led to their success in the certification process and that non-certified teachers similarly persist with their preferred approaches. Whether these differences in classroom practice are of importance with regard to objectively measured student achievement was not established.

NBPTS Validity: Value-Added Studies

Under the Bush administration, the U. S. Department of Education has been increasingly reluctant to fund NBPTS without better documentation of its effectiveness. In addition, states have become increasingly concerned with the cost of the bonuses and pay increases that teachers earn by becoming NBPTS-certified.

The studies discussed below address these issues by examining the relationship between NBPTS certification and improved student achievement. Their findings provide direct evidence of whether NBPTS certification is useful as an indicator of teacher quality and whether value-added achievement gains would be an even more useful teacher quality indicator.

Considering the divergence between NBPTS's objectives and that of public policy, there is a certain irony to the recent spate of reports. With the exception of Stone's (2002) report on NBPTS-certified teachers in Tennessee, these studies endeavor to show that the NBPTS standards lead to greater student gains on standardized tests. They do so even though the underlying NBPTS principles represent a different set of educational values and priorities.

Be that as it may, the studies synthesized below are important both because of what they reveal about NBPTS certification and because of what they suggest about the advantages of value-added assessment as an alternative to NBPTS certification. All are available online:

- Stone, J. E. (2002, May 1). *The value-added achievement gains of NBPTS-certified teachers in Tennessee: A brief report*. College of Education, East Tennessee State University. Retrieved January 18, 2005 from Education Consumers ClearingHouse: http://www.education-consumers.com/briefs/stoneNBPTS.shtm

- Goldhaber, D. and Anthony, E. (2004, April 27). *Can teacher quality be effectively assessed?* Urban Institute. Retrieved January 18, 2005 from http://www.crpe.org/workingpapers/pdf/NBPTSquality_report.pdf
- Vandevoort, L., Amrein-Beardsley, A., and Berliner, D. (2004, September 8). National Board certified teachers and their students' achievement." *Education Policy Analysis Archives, 12,* (46). Retreived January 18, 2005 from http://epaa.asu.edu/epaa/v12n46/v12n46.pdf
- Cavalluzzo, L. (2004, November). *Is National Board certification an effective signal of teacher quality?* CNA Corporation. Retrieved January 18, 2005 from http://www.cna.org/documents/CavaluzzoStudy.pdf

Each relies on one or another form of value-added achievement gain as the criterion against which the validity of NBPTS certification is judged. Each attempts to answer the question of whether students taught by Board-certified teachers exhibit demonstrably greater achievement gains than those taught by non-certified teachers.

What they found was that NBPTS-certified teachers produced achievement gains that were only slightly larger than those produced by non-certified colleagues, i.e. *effect sizes* in the neighborhood of eight percent of one standard deviation of the post-test scores.

That they found only small differences raises the primary question to which the present analysis is directed: Is NBPTS certification is worth the effort and expense that it entails?

It also raises the question of whether a certification program that can result in the allotment of monetary awards to low-performing teachers is acceptable from a policy standpoint. If the average performances of NBPTS-certified and non-NBPTS-certified teachers differ by only a small amount, it is virtually inevitable that a substantial number of the certified group are performing below the average of the non-certified group. It is equally inevitable that many members of the non-certified group are outperforming the certified group.

Beyond the matter of whether NBPTS certification is an adequate proxy for effective teaching is the second major question addressed by the present report: Might teacher quality be more accurately, conveniently, and inexpensively assessed by direct examination of value-added gains?

The studies examined below use various forms of value-added achievement gain as the criterion for teacher performance. Implicit in this alignment of variables is recognition of achievement gain as a direct and compelling measure of the ends sought by public policy. At its best, NBPTS certification is only a fallible estimate of this criterion.

In addition to its inherent superiority, the value-added alternative appears to have other important advantages:

1. The use of value-added assessment would largely eliminate the issues of misidentified teachers and misdirected bonuses. Teachers identified as high-performing on the basis of value-

added gain would, by definition, be high performers in the area of greatest concern. If local officials desired to include a broader range of considerations in teacher assessments, the added factors could be weighted and combined with achievement gain to produce an overall evaluation.

2. The effort and expense required to identify the best teachers would be substantially less than that required by the NBPTS. NBPTS assessment now costs $2300, and teachers have to undergo a laborious and time-consuming assessment process. Value-added assessment costs only a small fraction of that amount and requires none of the teacher time and effort.

3. With value-added assessment, teaching excellence can be assessed on the basis of annual job performance instead of a once-every-ten-years assessment of qualifications. In effect, value-added data permits judgments to be based on the realization of excellence, not just the potential for excellence. It is the realization of excellence that benefits students.

Given these several considerations, our analysis is focused on two main questions: 1) Is NBPTS certification a suitable proxy for student achievement gain, and 2) would the aim of improved student achievement be better served by a reliance on value-added achievement data.

Stone

In 2002, J. E. Stone at East Tennessee State University published a brief report on the classroom effectiveness of the 16 NBPTS-certified teachers for whom "teacher-effect" data was available in Tennessee's Value Added Assessment System (TVAAS). He found that none of them produced the achievement gains necessary to earn recognition as an exceptional teacher either by the state's criterion or by the criterion used in one of Tennessee's largest school districts.[1]

The teacher-effect data analyzed by Stone was comprised of 123 teacher-by-subject-by-year achievement gain scores. Of this group, 18 scores reached or exceeded the criterion necessary to be regarded as "exceptional" (115% of the school district average) and 13 were substantially below average (85% of the district average).

Stone's findings are of particular significance because TVAAS uses the highly sophisticated "mixed model" statistical analysis developed by Dr. William Sanders. Sanders' methodology is gaining popularity among the states, and its teacher-effect scores are regarded as far more accurate than the simple pretest to posttest comparisons used in other studies of achievement gain.

As the first empirical report to raise questions about the validity of NBPTS certification, Stone's study was the subject of extraordinary criticism--mostly coming from NBPTS stakeholders. A panel appointed by Ted Sanders, President of the Education Commission of the States (ECS), and led by Susan Fuhrman, dean of the University of Pennsylvania Graduate School of Education, flatly dismissed Stone's call for suspension of funding until NBPTS

certifications could be independently validated. The panel included Dominic Brewer, director of education at the RAND Corporation; Robert Linn, professor of education at the University of Colorado at Boulder and co-director of the National Center for Research on Evaluation, Standards and Student Testing; and Ana Maria Villegas, professor of curriculum and teaching at Montclair State University.

It now appears that the panel's conclusion may have been premature, if not misleading. Although the three subsequent studies found statistically significant differences between NBPTS-certified teachers and their peers—a matter discussed below—their results have proven entirely consistent with the small but inconsequential differences found by Stone. The primary distinction between Stone's findings and that of the other three studies is that Stone explicitly concluded that the NBPTS teachers were not exceptional and that the program should not be expanded until the worth of NBPTS certification could be independently validated.

ECS has not undertaken similar assessments of any subsequent reports, however, controversy about the value of NBPTS certification and the role of conflict of interest in validation research has continued to reverberate (Stone, 2003).

Stone's report discussed only the number and percentage of teacher-effect scores in reference to local school district averages. However, the data provided in his Appendix A make it possible to determine that the median teacher-effect for NBPTS-certified teachers was 102% of the local averages, i.e., a 2% advantage for NBPTS-certified teachers relative to their local peers.

Such a small gain is consistent with the findings of the other three studies considered in this analysis, and it may be statistically significant. As demonstrated by the following estimate, however, it is very much smaller than the teacher-effect average that would have been attained had each district's exceptional teachers been selected on the basis of their value added performance. For example, had a group of teachers been selected on the basis of whether their value-added gains reached the "A" level criterion used by Tennessee (115%), their average achievement gain would have been at least 7 times greater than the NBPTS group.

Goldhaber and Anthony

Dan Goldhaber of the University of Washington and Emily Anthony of the Urban Institute (2004) compared 303 of North Carolina's NBPTS-certified teachers to their non-certified peers on student achievement in 3rd, 4th, and 5th grade reading and mathematics—a pool of nearly 400,000 students.

Three groups were compared: NBPTS certified teachers, unsuccessful NBPTS applicants, and non-applicants. The following table is reprinted from Table 1 (p. 34) of Goldhaber and Anthony (2004):

Student Test Scores						
Variable	Non-applicants		Applicants, Not NBPTS Certified		Applicant, NBPTS Certified	
	Reading	Math	Reading	Math	Reading	Math
Post-test	149.47 (9.94)	150.39 (12.34)	149.47 (9.72)	149.80 (12.93)	151.52 (9.72)	152.38 (12.29)
Pre-test	143.78 (10.19)	140.64 (12.80)	143.65 (10.28)	140.67 (13.26)	145.34 (10.35)	142.17 (13.01)
Growth in test score in one year	5.69 (6.13)	9.75 (6.92)	5.83 (6.27)	9.14 (6.64)	6.18 (6.37)	10.21 (7.00)

The students taught by NBPTS-certified teachers gained slightly more than those taught either by unsuccessful applicants or by non-applicants. The differences, however, were trivial relative to the achievement gap faced by the 15-20% of North Carolina 3rd graders who are most in need of good teaching, i.e., those whose reading and mathematics scores are at *Level I*. As described by the North Carolina Department of Public Instruction (1995), Level I students are behind their grade mates by 16.2 points in reading and 22.1 points in mathematics. Given differences of this magnitude and the slight advantage associated with NBPTS certification (.49 points/year in reading and .46 points/year in mathematics), decades would be required to close the achievement gap.

Significance

When differences between groups are small, researchers apply tests of statistical significance to determine whether they are an artifact of sampling error. Differences that prove unlikely to be the result of error are termed *statistically significant.*

Although the term is widely used and frequently misunderstood, researchers universally agree that statistical significance does not mean practical significance. Statistical significance is the minimum required for a result to be considered something more than a chance outcome. Practical significance is the matter of whether a result is of sufficient size to be useful, e.g., large enough to substantially reduce the learning gap.

A finding of statistical significance is heavily influenced by the size of the groups that are compared within a given study. When very large groups are analyzed, trivial differences can reach statistical significance—which is exactly what happened in the Goldhaber and Anthony study.

In the interest of avoiding any misunderstanding, the authors might have given more attention to this well known distinction (Leamer, 1983). Unfortunately, their statements about significance were broadly disseminated by the NBPTS and the Center on Reinventing Public Education resulting in the widespread acceptance of the notion that NBPTS certification has

been unequivocally validated. In truth, given the size of the groups studied by Goldhaber and Anthony, the challenge would have been to find differences that were *not* statistically significant.

Moreover, as a technical matter, the use of statistical significance tests may not have been appropriate given the nature of the Goldhaber and Anthony data. Significance pertains to the matter of drawing inferences about populations from samples. It can be argued that the pool of students analyzed by Goldhaber and Anthony was the relevant population.

Effect Size

Because statistical significance is not a valid indicator of practical significance, researchers typically report the *effect size* statistic. It provides a way to avoid the over-interpretation of small differences when large samples are used—a corrective that is much needed in the present case.

In a few instances, Goldhaber and Anthony do report effect sizes, yet the message that these outcomes are practically unimportant does not come through. For example, in describing the differences between non-certified (i.e., unsuccessful) applicants and Future NBCT teachers (i.e., a teacher who will eventually become NBPTS certified), the authors conclude that:

> The magnitudes of the Future NBCT coefficients suggest that student gains produced by the teachers who are certified by NBPTS exceed those of non-certified applicants by about 4 percent of a standard deviation in reading and 5 percent of a standard deviations in math (based on a standard deviation of 9.94 on the end-of-year reading tests and 12.34 on the end-of-year math tests). These effect sizes are of the same order of magnitude as those found for math teachers having a bachelor's degree in their subject area (Goldhaber and Brewer, 1997). (p. 14).

Only readers familiar with the surprising result of Goldhaber and Brewer's 1997 study would know that the earlier study characterized its observed effect sizes as "relatively small."

Reference to Jacob Cohen's guidelines for interpreting effect sizes would have been helpful. A recognized authority in the matter, Cohen (1969, 1988) refers to findings of the size described by Goldhaber and Anthony as *small* to *trivial*. According to Cohen, greater than 50 percent is "large," 30 to 50 percent is "moderate," and 10 to 30 percent is "small." Anything less than 10 percent is "insubstantial, trivial, or otherwise not worth worrying about."

That Goldhaber and Anthony's effect sizes are relatively inconsequential may be determined by comparing them to the effects produced by other educational interventions. Lipsey & Wilson (1993) reported an extensive list of effect sizes drawn from meta-analytic studies of educational programs and teaching methodologies. Their median size was 50 percent. Walberg (in press) compiled a similar list that shows, for example, an effect size of 125% for the use of *cues* in instruction. That finding along with many others reported by Walberg suggests that changes in teaching style are far more likely than NBPTS certification to increase student achievement.

Results

Did the findings of Goldhaber and Anthony's study demonstrate that NBPTS-certified teachers are more effective in bringing about student achievement gains? Given the small differences that were found, their findings essentially support the opposite conclusion—i.e., that NBPTS-certified teachers are virtually indistinguishable from their non-certified peers.

The overlap between the certified and non-certified groups was very substantial. Over 40% of the scores attributable to non-certified teachers were above the average of the NBPTS-certified group. Conversely, over 40% of scores attributable to the NBPTS-certified teachers were below the average of the non-certified group. It is doubtful that policymakers understood just how much NBPTS-certified teachers overlapped the non-certified group when they agreed to fund the various state and local salary awards for NBPTS certification.

Beyond the matter of misplaced awards, would North Carolina get more "bang for the buck" by dispensing with NBPTS certification and selecting exceptional teachers on the basis of their value added scores? Clearly, the answer is yes.

For example, given the data provided by Goldhaber and Anthony, it is possible to estimate the student achievement gains that were being produced by teachers whose students were in the top 10% of value-added gains, i.e., those teachers whose value-added gains equaled or exceeded the 90^{th} percentile (Z=+1.282) of the non-applicant group.[2]

Using this criterion and the means and standard deviations from the Student Test Scores table (see above), the following value-added gains may be estimated:

Reading: Mean = 5.69; Standard Deviation = 6.13

Estimated achievement gain of top 10% > 5.69 + (1.282 x 6.13) > 13.55

Math: Mean = 9.75; Standard Deviation = 6.92

Estimated achievement gain of top 10% > 9.75 + (1.282 x 6.92) > 18.62

The estimated reading gain of 13.55 points is the minimum score necessary for inclusion in the top 10% group. A gain of 18.62 points is the minimum for the top 10% of math teachers. The averages of the two groups would be higher.

These are very large gains relative to those produced by NBPTS-certified teachers. The estimated reading gain is 27 times as large as the .49 produced by the NBPTS-certified group and the estimated math gain is 40 times as large as the .46 produced by the NBPTS-certified teachers. Given that "effect size" is achievement gain divided by the standard deviation of the dependent variable, the gains in both subjects would be at least 128%--as compared to the 6-14% effect sizes found for NBPTS-certified teachers.

Vandevoort, Amrein-Beardsley, and Berliner

Published in *Education Policy Analysis Archives*, "National Board Certified Teachers and Their Students' Achievement" seems extraordinarily critical of the articles and studies that fail to support NBPTS certification. For example, the authors are incredulous that Podgursky (2001) would question the value of the 13 teaching dimensions that Bond, Smith, Baker, and Hattie (2000) examined in their early attempt to validate NBPTS (Vandevoort, Amrein-Beardsley, and Berliner, 2004, p. 14). Similarly, they respond to Ballou's assertion that Board's standards were vague by questioning his understanding of how professions develop their certification programs. Their review of the literature seems less a review than a defense of NBPTS.

Conversely, the authors express complete confidence in Goldhaber and Anthony's supportive findings:

> ". . . [Goldhaber and Anthony] 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." (p. 13)

Methodology

Vandevoort, Amrein-Beardsley, and Berliner invited all Arizona teachers certified by NBPTS in Early Childhood or Middle Childhood (N=67) to furnish their student test scores and to complete a series of questionnaires. Thirty-five (44%) cooperated.

Using scaled scores in reading, math, and language for 1999-2003, the authors compared the achievement gains of NBPTS-certified teachers to those of non-certified teachers in each of 48 year-by-grade-by-subject cells (Tables 1-4). On the average, the 48 differences were small—2.45 points on a scale of approximately 65 points (p. 34). In only 11 of the 48 (23%) comparisons did the NBPTS-certified teachers outperform their non-certified colleagues by a statistically significant amount.

Again, the cautions pertaining to interpretation of statistical significance are applicable. Specifically, "significant" as used by Vandevoort, Amrein-Beardsley, and Berliner means only that there was less than a 5% chance that an observed difference was the result of error.

Of the remaining 37 comparisons, there were 13 in which the non-certified teachers produced greater gains than their NBPTS-certified colleagues; but the gains fell short of statistical significance.

For reasons not explained, the authors exclude these 13 comparisons as they discuss the results of their study. Rather than characterizing the 11 significant comparisons as 23% of the total of 48, they say the 11 were 31.4% of the 35 comparisons in which the NBPTS-certified teachers produced greater gains (p. 34). It would seem that the former would be the more accurate statement.

Also noteworthy was the absence of comment regarding the percentage of comparisons that did not reach significance. Of the 48 comparisons between NBPTS-certified and non-certified teachers, 75% were too small to be considered statistically significant. The obvious but unstated conclusion would seem to be that NBPTS certification in most cases made no statistically discernable difference.

Perhaps more importantly, it appears that Vandevoort, Amrein-Beardsley, and Berliner's finding of 11 statistically significant comparisons may have been a substantial overestimate. When multiple tests of significance are employed, the result is ordinarily corrected for the likelihood of false positives. Applying the Bonferroni correction for multiple comparisons, it appears that only 2 of the 48 (4%) should have been considered significant. Less conservative multiple-comparison tests could have been used but all would have reduced the number of significant differences.

As was true with the Goldhaber and Anthony study, Vandevoort, Amrein-Beardsley, and Berliner report effect sizes for each of the 48 comparisons (Tables 1-4). Although the authors characterize them as "of considerable importance," they average 12%, which is within Cohen's "small" range (p. 34). Too, they are small with respect to the achievement gaps that NBPTS-certified teachers are said uniquely qualified to remediate. For example, the average difference in reading achievement gain reported in Tables 1-4 is 2.39 points. By contrast, the SAT-9 scale score gap between students with "Limited English Proficiency" and "English Proficient" students in California for grades 2-6 is in the 15-30 point range (Thompson, DiCerbo, Mahoney, & MacSwan, 2002).

Comparing the average effect size of 12% found by Vandevoort, Amrein-Beardsley, and Berliner to the 128% effect size that would have been produced by the teachers who create the top 10% of value added gains again illustrates the substantial advantages that could be derived from identifying teachers by the latter means. In reading, for example, the advantage that would be expected from the selection of NBPTS-certified teachers would be 13% multiplied times the standard deviation of 23.23 = 3.02 points of gain (p 35). By comparison, the advantage that would be expected from the selection of teachers in the top 10% of value-added gains would be a minimum of 128% x 23.23 = 29.73 points of gain—nearly 10 times as much. The same kind of comparative advantage would be found in math (14% versus 128%) and language (7.5% versus 128%).

Contrary to their conclusion that the effect of NBPTS-certified teachers was "not trivial" (p.34), the statistical findings reported by Vandevoort, Amrein-Beardsley, and Berliner indicate a small effect—clearly, one far smaller than would be expected of teachers selected on the basis of a top-10% value-added performance.

Cavalluzzo

The most recent NBPTS certification study (Cavalluzzo, 2004) examined the performance of students taught by NBPTS-certified teachers in the Miami-Dade County Public Schools. FCAT mathematics scores from over 100,000 ninth and tenth grade students were used to compare the performance of 61 NBPTS-certified teachers to their non-certified peers.

Like Goldhaber and Anthony (2004) and Vandevoort, Amrein-Beardsley, and Berliner (2004), Cavalluzzo found a statistically significant but small difference: "Although NBC teachers have higher post-test scores than other groups, their students' gains differ little from those of other teachers" (p. 17).

Cavalluzzo's Table 2 (p. 18) indicates that the students taught by the 61 NBPTS-certified teachers gained 66.70 points per year while the students taught by the 1947 teachers who were *not involved* with NBPTS gained 65.45 points--a 1.25 point difference.

No matter what kind of statistical analysis is applied to this difference, it is trivial relative to the magnitude of the achievement gains necessary to bring low performing students up to minimally acceptable levels of performance. Approximately 35% of Miami-Dade 10th grade math students perform at Level 1 on the developmental achievement scale of the five-level Florida Comprehensive Achievement Test (FCAT; Department of Research Services, 2002). Level 1 scores range from 1068-1831. The current minimum passing score is 1889 and 1947 is the minimum Level 3 score (Florida Department of Education, 2005). In other words, the highest performing Level 1 students are 58 points below the minimum necessary to pass and 116 points below the minimum for basic math competence.

Using regression analysis, Cavalluzzo (2004, Table 3, p. 27) estimated that NBPTS-certified teachers were responsible for a 7.4% greater increase in achievement than were "otherwise similar teachers." Although this estimate may be inflated as a result of her approach to analyzing the data, it is still far smaller than the 128% effect size that would be obtained by selecting teachers from the top 10% if value-added gains.

Summary and Conclusion

For over a decade, questions have persisted regarding the worth of NBPTS certification. Thus far, studies have shown that it is more clearly an indicator of teacher commitment to NBPTS's Five Core Propositions (National Board for Professional Teaching Standards, 2005) than it is a predictor of improved student achievement.

The underlying problem seems to be that NBPTS's favored teaching style not well suited to the realization of the public's primary policy objective: improvement in objectively measured student achievement.

Beginning with Stone's 2002 report on the value-added achievement gains of NBPTS-certified teachers in Tennessee, a series of studies have sought to determine whether NBPTS-certified teachers are exceptionally effective in the classroom.

Stone found median gains that were approximately 2% above local school district averages for the 16 NBPTS-certified teachers for whom data was available—well short of the 15% gain that is considered exceptional in Tennessee's accountability system.

Goldhaber & Anthony (2004) found similar small gains for 303 NBPTS-certified teachers in North Carolina. The differences between the certified and non-certified teachers were found to be statistically significant. However, their practical significance remains in question.

The study of 35 NBPTS-certified teachers in Arizona by Vandevoort, Amrein-Beardsley, and Berliner (2004) found a small but statistically significant advantage favoring the NBPTS-certified group. Again, however, practical importance was in doubt.

Cavalluzzo (2004) studied the performance of 61 NBPTS-certified teachers in the Miami-Dade County public schools found that "their students' gains differ little from those of other teachers." The difference was statistically significant but small in absolute size.

Taken together, the studies examined by present report show that the achievement gains associated with NBPTS certification are small and thus give rise to the question of whether its costs are worth its benefits.

The costs are substantial. North Carolina, for example, has approximately 32,400 third, fourth, and fifth grade teachers. At $2,300 per teacher, the cost of NBPTS assessment for 10% this group would be $7.4 million.

The monetary expenses are not the only costs. Teachers who have completed NBPTS certification report that the process requires 150-200 hours of uncompensated time (Task Force on Teaching and Student Achievement, n.d., p. 40)—an estimated $3,000 to $4,000 at $20/hour.

By contrast, value-added assessment of teacher effectiveness is only a fraction of the cost with an enormous advantage in accuracy and convenience. For approximately $1/student and $25/teacher, available student test data can be analyzed to produce not only a teacher effectiveness indicator, but district, school, and individual student performance measures as well (Evergreen Freedom Foundation, 2001).

A policy of awarding bonuses to teachers who meet or exceed the 90th percentile of local district gains would not only eliminate the awkward possibility of misdirected awards, it would insure that all teachers who were doing truly exceptional work are rewarded for their talents and effort.

The effect sizes for NBPTS certification found by the above discussed studies are in the 6-14% range. In contrast, the 90th percentile is equivalent to an effect size of 128%. In other words, as a group, teachers selected on the basis of performance at or above the 90th percentile

would be responsible for an effect that is 10 to 20 times as large as the effect associated with NBPTS certification.

One last advantage of identifying exceptional teachers on the basis of value-added performance should again be noted. NBPTS certification and all other testing, licensure, and certification schemes attempt to measure attributes and qualifications that are said to be predictive of classroom performance. Value-added assessment, by contrast, is a direct measure of classroom performance and, as such, reflects that which is accomplished, not merely that which is promised.

In summary, a policy of rewarding teacher performance on the basis of value-added criteria would appear to be a far better choice than one of paying teachers for attaining NBPTS certification. A certification that would assure teacher quality remains a possibility, but one that exists more in theory than in practice.

References

Adamson, B. S. and White, W. E. (2001). *National board certified teachers: Questions and answers.* Comptroller of the Treasury, State of Tennessee. March. Retrieved January 18, 2005 from http://www.comptroller.state.tn.us/orea/reports/nbpts.pdf

Ballou, D. (2003). Certifying accomplished teachers: A critical look at the National Board for Professional Teaching Standards. *Peabody Journal of Education 78*(4), 201-219.

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

Buday, M. C. and Kelly, J. A. (1996). National Board certification and the teaching profession's commitment to quality assurance. *Phi Delta Kappan, 78*(3), 215-219.

Casas, M. (2003). The use of standardized tests in assessing authentic learning. *Teachers College Record*, ID Number: 11211. Retrieved January 18, 2005, from http://www.tcrecord.org/Content.asp?ContentID=11211

Cavalluzzo, L. (2004). *Is National Board Certification an Effective Signal of Teacher Quality?* Alexandria, VA: The CNA Corporation. Retrieved January 18, 2005 from http://www.cna.org/documents/CavaluzzoStudy.pdf

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

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

Department of Research Services, Office of Evaluation and Research, Miami-Dade County Public Schools. (2002). FCAT Performance and the Achievement Gap. *Research Brief, Vol. 020502*. Retrieved January 25, 2005 from http://drs.dadeschools.net/ResearchBriefs/ResearchBrief_01-02_502B.pdf

Evergreen Freedom Foundation. (2001). Value added assessment. *School Directors Handbook, VA-1*. Retrieved January 30, 2005 from http://www.effwa.org/pdfs/Value-Added.pdf.

Florida Department of Education (2005). *Understanding FCAT reports.* Retrieved January 21, 2005 from http://www.firn.edu/doe/sas/fcat/fcrpdes.htm.

Goldhaber, D. and Anthony, E. (2004). *Can teacher quality be effectively assessed?* Washington, DC: Urban Institute. Retrieved January 18, 2005 from http://www.urban.org/UploadedPDF/410958_NBPTSOutcomes.pdf

Goldhaber, D. D. and Brewer, D. J. (1997). Evaluating the effect of teacher degree level on educational performance. In J. William Fowler (Ed.), *Developments in School Finance 1996* (pp. 197-210). Washington, DC: National Center for Education Statistics. Retrieved February 6, 2005 from http://nces.ed.gov/pubs97/97535/97535l.asp

Holland, R. and Soifer, D. (2004). *Good ideas: Six valuable state and local education reforms.* Arlington, VA: Lexington Institute. Retrieved January 21, 2005 from http://www.lexingtoninstitute.org/education/pdf/GoodIdeas.pdf.

Kraft, N. (2001). *Standards in higher education: A critical analysis of NCATE, INTASC, and NBPTS.* Paper presented at the American Educational Research Association annual meeting, Seattle, WA.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment, confirmation from meta-analysis. *American Psychologist, 48*(12), 1181-1209.

Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review*, 73(1), p. 31-43.

National Board for Professional Teaching Standards. (2005). *What Teachers Should Know and Be Able to Do: The Five Core Propositions of the National Board.* Retrieved January 18, 2005 from http://www.nbpts.org/about/coreprops.cfm

North Carolina Department of Public Instruction, Office of Accountability (1995). *Research on end-of-grade testing: Characteristics of Level I students.* Retrieved January 18, 2005 from http://www.dpi.state.nc.us/accountability/testing/briefs/archives/V1N8.HTM

Podgursky. M. (2001). Defrocking the national board: Will the imprimatur of "Board Certification" professionalize teaching? *Education Matters,* Summer. Retrieved January 18, 2005 from http://www.educationnext.org/20012/79.pdf

Sack, J. (2003, November 12). Board stamp for teachers raising flags. *Education Week*, p. 1, 18. Retrieved January 18, 2005 from http://www.edweek.org/ew/articles/2003/11/12/11board.h23.html

Smerdon, Becky A., Burkam, David T., and Lee, Valerie E. (1999). Access to constructivist and didactic teaching: Who gets it? Where is it practiced? *Teachers College Record, 101*(1), 5–34. Retrieved January 18, 2005 from http://www.tcrecord.org/PDF/10423.pdf

Stone, J. E. (2003, June 6). Buyers and sellers of educational research. *Chronicle of Higher Education*, p. B12. Retrieved January 18, 2005 from http://chronicle.com/free/v49/i39/39b01202.htm

Stone, J. (2002). *The value-added achievement gains of NBPTS-certified teachers in Tennessee: A brief report.* College of Education, East Tennessee State University. Retrieved January 18, 2005 from http://www.education-consumers.com/briefs/stoneNBPTS.shtm

Task Force on Teaching and Student Achievement. (n.d.). *Teaching and learning: Meeting the challenge of high standards in Alabama.* Montgomery, AL: A+ Education Foundation. Retrieved February 2, 2005 from http://www.bestpracticescenter.org/publ/index.html.

Thompson, M.S., DiCerbo, K.E., Mahoney, K. and MacSwan, J. (2002, January 25). ¿Exito en California? A validity critique of language program evaluations and analysis of English learner test scores. *Education Policy Analysis Archives, 10*(7). Retrieved January 18, 2005 from http://epaa.asu.edu/epaa/v10n7/.

Vandevoort, L., Amrein-Beardsley, A., and Berliner, D. (2004). National Board certified teachers and their student achievement." *Education Policy Analysis Archives, 12*, 46. Retrieved January 18, 2005 from http://epaa.asu.edu/epaa/v12n46/v12n46.pdf

Walberg, H. J. (in press). Improving educational productivity: An assessment of extant research. In Rena Subotnik and Herbert J. Walberg (Eds.), *The Scientific Basis of Educational Productivity*. Greenwich, CT: Information Age Press.

---

[1] Chattanooga, Tennessee awards merit bonuses to teachers whose value-added achievement gains exceed 115%of the local district average. The program has substantially boosted achievement in the effected schools (Holland & Soifer, 2004).

[2] The NBPTS estimates that the top 10% of teachers will eventually be certified (Adamson & White, 2001).