

Value-Added Assessment and National Board for Professional Teaching Standards (NBPTS) Certification

Dr. George K. Cunningham
University of Louisville

John Stone
East Tennessee State University

Introduction

- The purpose of this presentation is to describe a particular application of value added principles: the way they have been used to assess the validity of the National Board of Professional Teacher Standards certification program.
 - An obvious way to validate the NBPTS process is to determine whether students taught by certified teachers show greater improvement in academic achievement than those who are not certified.
-

Two divergent perspectives on education reform

- ❑ NBPTS is based on the belief that the quality of a teacher is not to be found in the academic achievement of his or her students.
 - ❑ With NBPTS, teacher quality is much more closely associated with teacher beliefs and dispositions.
-

Related organizations

- NBPTS is part of a set of closely related organizations.
 - National Commission on Teaching and America's Future (NCTAF)
 - National Council for the Accreditation of Teacher Education (NCATE)
 - Interstate New Teacher Assessment and Support Consortium (INTASC).
-

Philosophy of the four related organizations

- ❑ Instead of academic achievement, these organizations advocate the importance of “learning.” By this they seem to mean the activity, actions, and dispositions of teachers rather than the academic achievement of students.
 - ❑ These organizations question the importance of academic achievement and express outright hostility to educational reform based on standards and accountability that every state has adopted.
-

Standards-based education reform

- The educational reform advocated by governors and state legislatures, which is strongly supported by the public, asserts that the best way to judge a principal, school, or teacher is by the academic achievement of the students in the school.
 - Academic achievement is operationally defined as performance on academic achievement tests.
-

Top ten states in terms of certified teachers and their yearly teacher bonuses for certification

□ 1. North Carolina	6,641	12%
□ 2. Florida	4,940	10%
□ 3. South Carolina	3,225	\$7,500
□ 4. California	2,664	\$10,000+
□ 5. Ohio	2,172	\$2,500
□ 6. Mississippi	1,761	\$6,000
□ 7. Georgia	1,321	10%
□ 8. Oklahoma	858	\$5,000
□ 9. Illinois	824	\$3,000
□ 10. Alabama,	632	\$3,000

Costs of NBPTS Certification

- ❑ Some states are backing away from rewarding certified teachers because of the escalating expense.
 - ❑ When the number of certified teachers gets too large, the cost becomes prohibitive.
 - ❑ The U.S. Dept. of Ed. no longer pays for NBPTS
-

Certification Process

- Four portfolio exercises
 - Samples of student work.
 - Videotapes of teaching
 - Self reflective commentary
 - Evidence of teacher community service and work with parents.
 - A response to 4 essay questions conducted at Sylvan Learning center
-

Certification Process

- ❑ Registration costs \$2300.
 - ❑ Most states and many school boards provide this fee for the candidate or they loan the money and cancel the loan if the teacher is certified.
 - ❑ The overall pass rate is about a 50 percent. For African American candidates the pass rate is slightly above 15 percent.
 - ❑ Most certified teachers are upper middle-class females who generally teach in advantaged schools.
-

Validity of NBPTS

- ❑ The underlying philosophy of NBPTS, like NCTAF, NCATE, and INTASC is hostile to standardized achievement testing.
 - ❑ Historically the certification has been validated with surveys of teachers and administrators with a heavy reliance on anecdotes.
-

Hostility to achievement testing

- The level of hostility to standardized achievement testing can be appreciated in this quote from a validation of teacher certification sponsored by NBPTS and conducted by Bond et. al. (2000). The reference can be seen in a following frame.
 - “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”
-

Need for validity studies

- ❑ Many states are paying a large amount of money to support this certification.
 - ❑ At the same time, these states have adopted educational reform accountability systems that utilize standardized achievement tests.
 - ❑ They have assumed that the quality of certified teachers should be reflected in better academic achievement.
 - ❑ This has led to demands for better validations of NBPTS certification based on the assessments of student achievement.
-

NBPTS Validity studies

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

NBPTS Validity studies

- Goldhaber, D. and Anthony, E. (2004). NBPTS certification: Who applies and what factors are associated with success. Urban Institute.
http://www.evansuw.org/FAC/Goldhaber/pdf/NBPTS_A-S.pdf
 - Vandervoort, A; Amrein-Beardsley, A; and Berliner, D. (2004) National Board Certified Teachers and Their student achievement. *Education Policy Analysis Archives. Vol. 12, No. 46.*
-

NBPTS Validity studies

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

The Bond study

- ❑ The Bond study was an early attempt to demonstrate the validity of the certification process.
 - ❑ Reports in *Education Weekly* made it sound as though this study answered all questions about the validity of NBPTS.
 - ❑ It was a lengthy and expensive three hundred page study.
 - ❑ The study itself consisted of thirty-one certified teachers that were compared with 34 teachers who applied but failed to be certified.
-

The Bond study

- Instead of trying to ensure the comparability of the two groups, the 31 certified teachers were selected because they had high scores and the 34 of those not certified were selected because they had low scores.
 - This was done in order:
 - “To ensure that dependable differences between National Board Certified teachers and non-Certified teachers were detected...(Bond, et. al., 2000, p. 69).”
-

The Bond study

- ❑ Rejecting objective measures of student achievement, the authors chose instead to observe the teachers and survey students to learn from them whether their teachers exhibited teaching behaviors consistent with thirteen principles of good teaching, identified in a review of the literature.
 - ❑ Student achievement was assessed with a portfolio of student work and student responses to writing prompts.
 - ❑ The thirteen dimensions assessed whether teachers who were certified displayed the sort of learner-centered teaching behaviors treasured by the NBPTS.
 - ❑ Not surprisingly they did.
-

The Bond study

- ❑ NBPTS has defined good teaching as learner-centered instruction.
 - ❑ Certified teachers, by definition teach this way, otherwise, they would not have been certified in the first place.
 - ❑ When you compare certified teachers with those who failed the certification process, the certified teachers of course demonstrate that they embrace the methods and philosophy that got them certified.
 - ❑ At the same time, those denied certification do not demonstrate these behaviors and attitudes to the same degree.
-

The Bond study

- ❑ The authors of the study come to the unsurprising conclusion that certified teachers display the behaviors required for certification more consistently than those denied certification.
 - ❑ They then conclude that this proves that certified teachers are better.
 - ❑ This was more a reliability than a validity study.
-

Measures of achievement in the Bond study

- 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.
 - A study by Goldhaber and Anthony (2003) has established that students taught by NBPTS-certified teachers tend to be socio-economically advantaged and high achievers. They also teach in higher performing schools.
 - There were no differences in the writing performance of the students taught by the two teacher groups.
-

The Stone study

- ❑ This is the only study of the five that was wholly independent from NBPTS.
 - ❑ It failed to find differences between certified and non-certified teachers.
 - ❑ This study was attacked by supporters of NBPTS certification.
 - ❑ The attacks on this study were out of proportion with the modesty of the study.
 - ❑ The study used data from the Tennessee Value Added Assessment System (TVAAS).
-

The Stone study

- Tennessee only had 41 NBPTS teachers and only 16 of these taught in grades 3-8 and thus had value-added data available.
 - Using the criterion that Tennessee uses to classify student achievement gains, Stone found only 18 of the 123 teacher/subject/year teacher-effect scores for the NBPTS-certified teachers to be “exceptional,” i.e., at or above 115% of the gains produced by other teachers.
 - As importantly, he found 13 of the 123 scores were substantially below average, i.e., at or below 85% of the gains produced by other teachers.
-

The Stone study

- By comparing the performance of NBPTS-certified teachers to the merit pay standard used in an urban school district, Stone determined that none of the NBPTS-certified teachers would have qualified for a bonus.
 - in 16 out of the 16 available cases, NBPTS-certified teachers were not exceptional producers of student achievement.
-

Goldhaber and Anthony study

- ❑ The results of the Goldhaber and Anthony study were touted in the media as the resolution of all questions about the validity of NBPTS.
 - ❑ This study endeavored to validate NBPTS certification by showing that the students of certified teachers had higher academic achievement than those who were not certified.
-

Goldhaber and Anthony study

- ❑ This study presents a paradox because NBPTS is based on a philosophy that eschews standardized achievement testing.
 - ❑ Generally, supporters of NBPTS assert the benefits of this certification without reference to student achievement.
-

Goldhaber and Anthony study

- ❑ This is a very ambitious and large-scale study.
 - ❑ It includes 400,000 students, 303 teachers who were certified, and slightly more than 6,000 students taught by certified teachers.
 - ❑ This is a value-added study in that it uses gain scores computed by subtracting pre from post scores.
 - ❑ The value-added model employed lacks the sophistication of the Sander's model or approaches based on hierarchical linear modeling.
 - ❑ There would have been some advantages to using pre-test as a covariate as was done in the Vandevort and Cavaluzzo studies rather than gain scores.
 - ❑ These more complex models are also much more difficult to communicate to the public.
-

Goldhaber and Anthony study

- ❑ The authors compared the academic achievement of certified teachers with those who applied, but were not certified, and to teachers who did not apply for certification.
 - ❑ While the differences found were labeled significant, this assertion is misleading. With hundreds of thousands of subjects, inconsequential differences can be labeled statistically significant.
 - ❑ The authors could have concluded that the academic achievement of certified and uncertified teachers differed only by a trivial amount.
 - ❑ They chose instead to assert that certified teachers were more effective in increasing student academic achievement than non-certified teachers.
-

Goldhaber and Anthony study

- ❑ For all applicants, about 50 percent are certified.
 - ❑ In North Carolina, the ratio of African Americans to White applicants was 13 to 85 about the same as the teacher population in the state.
 - ❑ Only 15 percent of African Americans were certified.
-

Goldhaber and Anthony study

- ❑ The NBPTS certification process has one of the greatest adverse impacts of almost any assessment.
 - ❑ This is an interesting outcome for an organization that places great emphasis on the value of diversity.
 - ❑ Unlike other assessments that have had an adverse impact, NBPTS certification is not based on a cognitive test like the SAT.
 - ❑ It is instead based most heavily on beliefs and dispositions.
-

Methodological issues in the Goldhaber and Anthony study

- ❑ The authors fail to distinguish between statistical significance and practical significance or importance.
 - ❑ The number of subjects is so large that almost any difference would be statistical significant.
 - ❑ Inferential statistics are employed even though samples are not used to understand a population. The study includes the entire population of students in North Carolina.
-

Methodological issues in the Goldhaber and Anthony study

- ❑ Alpha levels are cited inconsistently.
 - ❑ Various statistical models are employed with the implication that only those that provide desired results are reported.
 - ❑ For example, the authors describe how they "experiment with using the Z-scores from various measures of teacher academic proficiency (p. 13)."
 - ❑ What they do not tell the reader is the criteria used for deciding which should be reported.
-

Methodological issues in the Goldhaber and Anthony study

- ❑ The authors report some effect sizes, but they do not do so systematically in tables and most appear only in footnotes.
 - ❑ Their interpretation of the effect sizes reported is misleading.
 - ❑ Most of the effect sizes reported would be considered insubstantial and trivial by Cohen, but Goldhaber and Anthony repeatedly cite the small effect sizes as indicative of meaningful differences.
-

□ The authors make the following assertion about the differences between non-certified applicants and Future NBCT teachers:

- “The magnitude 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).
-

Methodological issues in the Goldhaber and Anthony study

- 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); (Goldhaber and Anthony, 2004, p. 14)."
 - Unless the reader is familiar with the 1997 study cited, they would not know that the effect size in that study is characterized as small.
 - The suggestion that effect sizes of 4 and 5 percent are evidence for meaningful differences is stunning.
 - Citing the effect size of the possession of a bachelors degree in math, cited from a previous study is likely to mislead the reader.
-

Conclusions about the Goldhaber and Anthony study

- Despite the heroic number of subjects, the large number of variables, and the experimentation with different statistical models, in the end, despite all of the ambiguous uses of the term “significance,” they present no evidence that NBPTS certified teachers are any better than any other teachers.
-

Conclusions about the Goldhaber and Anthony study

- This begs the question of why North Carolina or any other state is willing to pay many millions of dollars on a program that primarily benefits White, middle class, female teachers and has little if any positive effect on student achievement.
-

Conclusions about the the Goldhaber and Anthony study

- The study seems to show that NBPTS-certified teachers are, practically speaking, indistinguishable from other teachers with regard to effectiveness.
 - The overlap between the groups is enormous. As can be determined from the Goldhaber & Anthony data, over 40% of non-certified teachers are more effective than the average of the NBPTS-certified group. Conversely, over 40% of NBPTS-certified teachers are less effective than the average of non-certified teachers.
-

The Vandervoort et. al. study

- ❑ This study is similar to the Goldhaber and Anthony study. It differs by focusing on the comparison of certified with non-certified teachers.
 - ❑ Both the Bond and Goldhaber studies focus on the comparison between certified teachers and teachers who sought certification, but who were turned down
-

The Vandervoort et. al. study

- The first part of the paper is a defense of NBCTS that strongly endorses those studies that support it while criticizing any studies or articles that fail to support it.
 - The authors of the study are in full agreement with the basic assumptions of NBCTS, which is that good teachers can be identified by their adherence to a set of beliefs that can best be described as teacher-centered, progressive, and constructivist.
-

The Vandervoort et. al. study

- They assert that studies that support NBPTS are beyond reproach. In responding to Podgursky's (2001) criticisms of the set of 13 dimensions of teaching expertise, Vandervoort et. al. are incredulous that anyone could question their merit.
-

The Vandervoort et. al. study

- They seem unaware that there could be educators who fail to believe that student-centered constructivist educational methods are the only acceptable approach.
 - The final statement on the issue and the one apparently intended to end all discussion, is in reference to Bond's response to criticism. It is as follows: "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."
 - Apparently only an apology by Bond would provide sufficient evidence for a flaw in his study.
-

The Vandervoort et. al. study

- Vandervoort's praise of the Goldhaber study is unstinting. They summarize by stating:
 - "These researchers believed that they (Goldhaber and Anthony) 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."
-

The Vandervoort et. al. study

- ❑ The authors identified 37 certified teachers who were willing to participate in the study, but only 34 completed the survey.
 - ❑ They were compared to a total of almost 60,000 students in the four grades being studied.
 - ❑ All students in grades 3 through 6 were included if they had complete data for the years being compared.
 - ❑ Scaled scores in reading, math, and language were examined for years 1999-2000, 2000-2001, 2001-2002, and 2002-2003. .
-

The Vandervoort et. al. study

- Gain scores were created by subtracting the first year scaled score from the second year scaled score.
 - These gain scores were used as the dependent variable in a general linear model design, which included NBCT status as the independent variable and, surprisingly, the first year scaled score as a covariate
-

The Vandervoort et. al. study

- ❑ They then reported the adjusted gains scores.
 - ❑ It is not clear how this design would work because in creating the gains scores, the effect of the first year score had already been removed.
 - ❑ To add it as a covariate would not explain any further variance.
-

The Vandervoort et. al. study

- ❑ There are 48 test of significance reported in this study (4 years by 4 grades by 3 subjects).
 - ❑ The results are reported in terms of the number of comparisons that favor NBCTs whether they are significant or not.
 - ❑ If differences are not significant, they are not different, and should not be reported.
 - ❑ When multiple tests of significance are reported, alpha slippage occurs.
-

The Vandervoort et. al. study

- ❑ It would be appropriate to apply the Bonferroni correction.
 - ❑ An examination of the p-values reported in the appendix indicates that there are only two comparisons that would be significant with this correction and a few others that are close.
-

The Vandervoort et. al. study

- ❑ The authors incorrectly interpret the effect scores they report.
 - ❑ An important purpose of effect scores is to provide a way to prevent over-interpretation of differences that are significant as the result of large samples.
 - ❑ Instead, Vandervoort et. al., like Goldhaber use effect scores as a way to convince the reader that the very small differences they have found are actually important.
-

The Vandervoort et. al. study

- ❑ The study includes a convoluted attempt to transform the effect scores into months or even weeks of grade equivalent gains.
 - ❑ Grade equivalents are a particularly imprecise derived score. They assume linearity where it is unlikely to exist and stray far from being interval data.
-

The Cavaluzzo study

- The Cavaluzzo study is similar to the Goldhaber and Vandevoot studies already discussed.
 - 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.
-

The Cavaluzzo study

- As was true with the previous two studies discussed, they found significant differences, which is not surprising given the size of the sample.
 - Cavaluzzo also improperly invokes effect sizes to make these slight differences seem meaningful.
-

Summary and conclusions

- ❑ For over a decade, questions have been asked about the validity of the NBPTS certification.
 - ❑ The problem is that this certification encourages a style of teaching that is not particularly effective in increasing student academic achievement.
-

Summary and conclusions

- There have been a series of validity studies, which have endeavored to establish that NBPTS certification has a positive effect on student achievement.
 - The best they can do is to claim statistical significance with heroically large sample sizes.
-

Summary and conclusions

- In the cases of the Goldhaber, Vandervoort, and Cavaluzzo studies, these results have been bolstered by the misleading use of effect sizes.
 - The costs of this certification is staggering.
 - In addition to the initial \$2300 application fee many states provide yearly bonuses.
-

Summary and conclusions

- South Carolina, a state that is not wealthy is paying \$24,187,500 a year in bonuses.
 - This is a certification program that primarily benefits well-off, white females teaching in advantaged schools.
 - If states want to reward successful teachers they could reward those teachers whose students show increases in standardized achievement performance.
-

Summary and conclusions

- ❑ Objections to such merit pay proposals usually focus on the unreliability and unfairness of assessments based on test scores.
 - ❑ Compared with the NBPTS process such a value-added process would be the very model of reliability, validity, and fairness.
-