

[Hoover Home](#) | [News](#) | [Get Involved](#) | [Search](#)

[About Hoover](#) | [Library & Archives](#) | [Research](#) | [Fellows](#) | [Publications & Outreach](#)

**EDUCATION**  
**next**

## Defrocking the National Board

by MICHAEL PODGURSKY

[home](#)

[unabridged](#)

[past issues](#)

[search](#)

[about](#)

[single issues](#)

[subscriptions](#)

### **The Certification System of the National Board for Professional Teaching Standards: A Construct and Consequential Validity Study**

**by Lloyd Bond, Richard Jaeger, Tracy Smith, & John Hattie**

*Center for Educational Research and Evaluation, University of North Carolina at Greensboro, October 2000*

The National Board for Professional Teaching Standards, through its series of comprehensive performance assessments of teaching proficiency, is identifying and certifying teachers who are producing students who differ in profound and important ways from those taught by less proficient teachers.” So concludes this National Board–selected group of researchers in their study comparing National Board–certified teachers with teachers who were unsuccessful in their bid for certification.

Although the federal government, states, school districts, and private foundations already have invested nearly \$200 million in producing and rewarding National Board–certified teachers, this is the first study assessing whether the National Board has actually succeeded in identifying “expert” or “master” teachers who perform better than their uncertified peers. The National Board and its supporters greeted the results with pleasure. “This study tells parents and the community, educators and policymakers that National Board Certification is a distinction that really matters,” National Board president Betty Castor declared in a press release. Lee Shulman, president of the Carnegie Foundation for the Advancement of Teaching, claims, “In no other profession will you find an equivalent study with these kinds of rigorous assessments . . . no comparison can be found.”

Unfortunately, there is much less in this report than the press releases imply. In effect, the report really tells us only that teachers who were certified by the National Board were more likely to display the types of behaviors the National Board favors. Such a circular exercise does not necessarily prove that National Board–certified teachers do a better job of raising student achievement. After \$500,000 (the cost of this U.S. Department of Education–funded study) and three years of research, the fundamental question remains unanswered: Is the National Board’s certification process a valid and cost-effective way of identifying the nation’s best teachers?

## A Rising Tab

The National Board for Professional Teaching Standards was created in 1987 on the recommendation of the Carnegie Corporation's Task Force on Teaching as a Profession. The board's mandate was to raise the professional status of teachers (and the quality of teaching) by creating a means to identify and certify the most accomplished teachers. The National Board likes to compare itself to the medical specialty boards. All doctors are licensed by their states, but most go on to obtain advanced training and voluntary certification from one of the 24 medical specialty boards. The National Board sees itself as providing a similar form of advanced certification, a signal of expertise and excellence.

The Carnegie Foundation provided the National Board's start-up funds, but, beginning in the early 1990s, the U.S. Department of Education became the National Board's primary source of support. In recent years the National Board has received roughly \$20 million annually from the U.S. Department of Education. Currently there are roughly 9,500 nationally certified teachers, with many more in the pipeline. Substantial pay increases now accompany board certification in many states and districts. The Los Angeles unified school district recently signed a contract with its American Federation of Teachers local that gives board-certified teachers a 15 percent bonus for the ten-year duration of a National Board certificate. Florida offers a 10 percent bonus for ten years plus an additional 10 percent if board-certified teachers agree to mentor other teachers. Ohio provides an annual bonus of \$2,500 for ten years. The Cincinnati teachers union negotiated an additional \$1,000 bonus, plus an additional \$4,500 if board-certified teachers serve as lead teachers.

In short, states and districts are beginning to incur substantial long-term costs in rewarding National Board teachers. If the National Board reaches its goal of having 105,000 certified teachers by 2006, states and districts may be spending nearly \$1 billion annually in additional compensation alone (not counting the \$2,300 National Board assessment fee and related costs). Moreover, with National Board teachers acting as mentors for new teachers, their influence will extend well beyond their numbers.

### (Not) Measuring Achievement

No study, however, has ever shown that National Board-certified

teachers are any better than other teachers at raising student achievement. Nothing has changed with the release of this report. The National Board's researchers rejected the use of student test scores as a measure of teacher performance, claiming, "It is not too much of an exaggeration to state that such measures have been cited as a cause of all of the nation's considerable problems in educating our youth. . . . It is in their uses as measures of individual teacher effectiveness and quality that such measures are particularly inappropriate."

To measure teacher quality, the authors began by trying to determine what good teachers know and do. This process of creating what the authors consider appropriate standards involved "a massive synthesis of meta-analyses that encompass over 200,000 research studies." These 200,000 studies included ethnographic along with conventional statistical studies. From this synthesis, the researchers claim to have distilled 13 principles of good teaching. Examples include:

- "Experienced expert teachers adopt a problem-solving stance to their work."
- "Experienced expert teachers aim at creating an optimal classroom climate for learning."
- "Experienced expert teachers are passionate about teaching and learning."

The researchers then developed methods for measuring and scoring such attributes in their sample of 65 teachers (31 who passed and 34 who failed their National Board assessment). For example, in order to assess the "multidimensional perception" of teachers (i.e., "Expert teachers develop a high level of 'withitness,' that is, they show that they are aware of events that occur simultaneously"), the researchers used a survey of each teacher's students and observed teachers during a three-hour, prearranged classroom visit. The goal of creating an "optimal classroom climate" was measured in a similar manner. On all 13 dimensions of teaching practice, National Board teachers outscored those who did not make the cut. In 11 of 13 cases, the differences were statistically significant.

**Although nearly \$200 million has been invested in producing and rewarding National Board—certified teachers, this is the first study assessing whether "board-certified" teachers perform better than their uncertified peers.**

The researchers also examined two measures of student performance: an "internal" example of student work provided by the teacher based on the lesson observed by the research team and an "external,

developmentally appropriate measure of writing proficiency.” The “internal” work samples obviously varied from teacher to teacher, but were graded according to a standard rubric. The students of the board-certified teachers scored significantly higher on the “internal” measure. No statistically significant differences were found on the “external” writing assignment. (Ten of the 65 teachers refused to provide student work, and a number of the classroom-based assignments were unscorable.)

There are numerous problems with this methodology. Let’s begin with the authors’ using 13 “dimensions of teaching expertise” as their evaluation criteria instead of student test scores. Even if we accept the dubious proposition that 200,000 studies provide a scientific basis for the authors’ 13 nebulous standards of good teacher practice, we can’t be sure that the ways in which the authors have chosen to measure these standards necessarily replicate those of the underlying studies. Exactly how did thousands of different studies, of varying methodological rigor, measure “an optimal classroom climate for learning”? The authors’ 13 dimensions of teaching practice are valid measurement criteria only if the authors can demonstrate that their measures exactly replicate those of the literature they cite. And if the underlying measure of student achievement in these studies was standardized tests, as was surely the case in many of them, why are such tests acceptable as measures of teacher quality in studies that are meta-analyzed and used indirectly, but unacceptable when they are used directly to assess teacher quality in a structured research design? Readers of this study simply have no way of knowing whether the researchers’ 13 measures of teacher expertise actually correlate with improved student achievement.

### Lack of Controls

When the authors actually did examine student performance, albeit with rather vague measures, they neglected to collect data that would have permitted them to adjust the performance data for students’ socioeconomic status, demographic characteristics, and previous levels of achievement. If the socioeconomic status and demographic characteristics of the classrooms taught by National Board teachers differ from those of noncertified teachers, measures of teacher quality that rely on student performance may be biased. Students of National Board teachers who exhibited superior academic performance may already have been performing at a high level when they entered class in the fall.

The authors acknowledge this limitation concerning their measures of student performance, but the issues of socioeconomic status and

previous student achievement are problems for all of the researchers' measures of teacher quality. For example, several of the 13 dimensions of teaching expertise were measured using student surveys, with questions such as, "An important reason why I do homework is because I like to learn new things," or, "I do my schoolwork because I'm interested in it." Students' family backgrounds and previous educational achievement are likely to influence their responses to such questions.

## Show Me the Money

*Although no studies show that National Board-certified teachers are any better than their peers, states and districts are rushing to reward them handsomely.*

Los Angeles	15% raise for 10 years
Florida	10% raise for 10 years, additional 10% for being a mentor teacher
Mississippi	\$6,000 bonus for 10 years
Louisiana	\$5,000 bonus for 10 years
Oklahoma	
Iowa	\$5,000 bonus for 10 years; \$2,500 if certified after May 1, 2000
Ohio	\$2,500 bonus for 10 years
Cincinnati	\$1,000 bonus in addition to Ohio bonus, plus \$4,500 for lead teachers

Students' family backgrounds are also likely to affect researchers' evaluations of teachers' classroom practice. Imagine two classrooms: one with well-behaved, highly motivated students from well-to-do families, the other with poorly behaved, unmotivated students from poor families. Now consider the scoring criteria. Under "Preventive and Reactive Classroom Environment," teachers receive the top score if they "provide effective management procedures with a comprehensive focus on student learning," but receive the lowest score if they "react to disciplinary incidents after the fact rather than trying to prevent them." On "Multidimensional Perception," teachers receive the top score if they "identif[y] events occurring simultaneously while maintaining a focus on instruction." A teacher who is "often overwhelmed by the complexity of classroom events" receives the lowest score. It is easy to imagine the bias introduced by differences in students across classrooms. If we took high-scoring,

“multidimensionally perceptive” teachers out of their well-to-do classrooms and put them in tough, low-income classrooms, they too might be “overwhelmed by the complexity of classroom events.”

In fact, there is evidence of significant socioeconomic differences between the classrooms of National Board teachers and those of unsuccessful certification candidates. In response to queries about this matter, the National Board provided some unpublished data. The 65 teachers were asked the following question: “Approximately what percent of the students in your class come from the following types of families?” Among the board-certified teachers, 44 percent reported that the largest share of their students came from “well-to-do families with few if any financial problems,” while only 8 percent reported that the largest share came from “families who cannot afford the basic necessities of food, clothing, and shelter.” Among unsuccessful candidates, however, the corresponding percentages were 21 percent in each category. By this measure, board-certified teachers were twice as likely to have children from wealthy families and less than half as likely to have poor children. This suggests that there may have been a major socioeconomic gap between the students of the two groups of teachers.

**If the National Board reaches its goal of having 105,000 certified teachers by 2006, states and districts may be spending nearly \$1 billion annually in additional compensation alone.**

Finally, there is the question of sampling. To determine whether the average National Board teacher is better than the average unsuccessful candidate, we need to draw random samples of both groups. The National Board’s researchers, however, chose a peculiar sampling scheme that oversampled particularly high-scoring National Board teachers and particularly low-scoring teachers who were unsuccessful. If we think of teachers who pass the National Board’s assessments as earning a C or better, this procedure amounted to oversampling teachers who earned As and Fs. The researchers are well aware of the effect of such sampling: “These groups were defined to ensure that dependable differences between National Board Teachers and non-Board Certified teachers could be detected.” In other words, they structured the sampling so as to increase the likelihood of finding an effect of National Board certification. This type of sampling is justified for some of the more complicated statistical analysis conducted in one section of the report. However, by sampling in this manner, the authors have rendered meaningless any simple comparisons of averages—the kinds of simple comparisons that are prominently displayed in the press release and the executive summary of their report. It is simply incorrect to claim that “NB teachers are superior in 11 of 13 dimensions” when the researchers have sampled

in this manner. We do not know whether simple random sampling would have yielded significantly higher means for National Board–certified teachers on 11 of 13 dimensions.

**Readers of this study simply have no way of knowing whether the researchers' 13 measures of teacher expertise actually correlate with improved student achievement.**

The resources available for this study (\$500,000, or roughly \$8,000 per teacher) would certainly have been more than enough to perform a rigorous analysis of the performance of National Board teachers vis-à-vis unsuccessful candidates, using a random sample of the two groups and adjusting for students' socioeconomic status and previous achievement levels. In fact, these resources probably would have been adequate to increase the sample to several hundred teachers. Such a study would ask not only whether the achievement scores of students of National Board teachers improved more than the scores of students of unsuccessful candidates, but also whether National Board certification was a cost-efficient way to identify excellent teachers. For example, would principal or parental evaluations have worked just as well? Or were less costly components of the teachers' National Board scores, such as the one-day assessment at a Sylvan Learning Center, just as effective as the costly, time-consuming (and coaching- or cheating-prone) portfolio in predicting student performance? The shortcomings of this study, the paucity of independent research on the National Board, and the large investments being made by states in rewarding National Board–certified teachers highlight the need for a rigorous and arm's-length cost-benefit study of National Board certification.

Published by the [Hoover Institution](#) © 2001 by the Board of Trustees of [Leland Stanford Junior University](#)

[home](#) | [past issues](#) | [search](#) | [about](#) | [subscriptions](#)