

# ECF Education Briefs

Issues in Public Education: Research and Analysis  
from the Education Consumers Foundation

June 2004 • v4n6

## Can Teacher Quality Be Effectively Assessed?

By Dan Goldhaber, University of Washington Center for Reinventing  
Public Education, Emily Anthony Urban Institute

Found online at [http://education-consumers.org/research/briefs\\_0604.htm](http://education-consumers.org/research/briefs_0604.htm)



**Education Consumers Foundation**

1655 North Fort Myer Drive, Suite 700 • Arlington, VA 22209  
[www.education-consumers.org](http://www.education-consumers.org)

# Can Teacher Quality be Fairly Assessed?

*Dan Goldhaber*

Teachers pay \$2,300 to be assessed by the National Board for Professional Teaching Standards (NBPTS). They earn pay increases up to \$7,500 per year if successful. To date, NBPTS has certified over 30,000 teachers.

Despite widespread reports to the contrary, the evidence brought forth by Goldhaber & Anthony does not strengthen the claim that NBPTS-certified teachers are substantially more effective than their colleagues in bringing about student achievement. If anything, the case for NBPTS certification may now be weaker.

Prior to Goldhaber & Anthony, only two small-scale investigations had examined the link between NBPTS certification and student achievement. Bond, Smith, Baker, & Hattie (2000) compared 31 NBPTS-certified teachers with 34 teachers who had applied for certification and failed. There was a slight edge in achievement favoring the NBPTS-certified group, however, the study was fraught with methodological difficulties (see Podgursky, 2002).

Stone (2002) examined the value-added achievement gains of the 16 NBPTS-certified teachers for whom data was available in Tennessee's state database. He found both above and below-average performance, and determined that none of the 16 would have qualified for the performance-based salary increase awarded by one of Tennessee's urban school districts. Stone's report was the first empirical study to raise doubt about the value of NBPTS certification; and as such, stirred considerable criticism.

## **Findings**

Goldhaber & Anthony's study compared all of North Carolina's 3rd, 4th, and 5th grade NBPTS-certified teachers to their colleagues around the state. It analyzed the achievement test scores of over 400,000 students taught by more than 16,000 teachers and concluded that NBPTS-certified teachers produced "significantly" greater gains in student achievement.

Its conclusion, however, referred only to the "statistical significance" of the results. In fact, the differences found by Goldhaber & Anthony were quite small and of doubtful scientific, educational, or practical importance (see Table 1, p. 29).

Less than one point (.49) in annual reading score gain separated NBPTS-certified teachers from nonapplicants (6.18 versus 5.69), and an even smaller difference (.35) separated them from those teachers who applied but were not selected (6.18 versus 5.83).

A similar pattern of small differences was observed with respect to the math scores. NBPTS-certified teachers differed from nonapplicants by less than one point (.46, 10.21 versus 9.75) and from teachers who applied but were not selected by slightly more than one point (1.07, 10.21 versus 9.14).

Statistically significant does not mean important

“Significance,” as the term is used in statistical reports, refers only to whether a numerical result is of sufficient size to be considered something more than the product of random error. It refers to a minimum or threshold level of outcome. It is not a synonym for “important.”

To their credit, the authors acknowledged that the observed differences were “. . . relatively small; the largest differential [being] in math between certified and non-certified teacher applicants, at just over a point on the exam or roughly 14% of a standard deviation in the growth of math scores”(p. 14).

What their comments did not make clear, however, is just how small these differences are relative to accepted benchmarks for “effect size.”

Effect size is the difference between two groups stated as a percentage of a standard deviation (i.e., “14%”). It is the appropriate statistic for gauging the importance of results such as those found by Goldhaber & Anthony.

Jacob Cohen (1969, 1988) --perhaps the foremost authority on effect size--offers the following guidelines: greater than 50% = “large,” 50-30% = “moderate,” 30-10% = “small,” and less than 10% = “insubstantial, trivial, or otherwise not worth worrying about.” Other authorities (Glass, McGaw, & Smith, 1981) caution that Cohen’s categories may be too liberal because they fail to take cost-effectiveness into account.

Goldhaber & Anthony found one effect size of 14% and three others in the 6-8% range. Clearly, effect sizes of this magnitude are of doubtful importance.

Another way of judging effect size is by comparison to the effects of other policies and interventions. Goldhaber and Anthony employ this practice when they compare their results to the effects of class-size reductions and to the effects associated with teachers’ holding a bachelor’s degree in their subject area (p. 18).

Mark Lipsey and David Wilson compiled a list of effect sizes drawn from hundreds of educational and psychological studies. The median was 34%.

Of particular relevance to Goldhaber & Anthony, Lipsey & Wilson found substantial differences in student achievement effects associated with various teaching methodologies. For example, the use of positive reinforcement produced an average effect of 117% whereas the “open classroom” approach to teaching produced effects ranging from plus 1% to minus 13%.

Plainly, such a range of effect sizes suggests that choice of teaching methodology has a far greater impact on student achievement than does NBPTS certification.

## **Conclusion**

Rather than demonstrate the superiority of NBPTS-certified teachers, Goldhaber & Anthony showed that NBPTS-certified teachers are nearly indistinguishable (see chart) from their colleagues.

Given that this same conclusion was suggested by Stone's earlier study, policies that reward teacher quality on the basis of NBPTS certification should be reconsidered. As shown by Goldhaber and Anthony and by Stone, statistical analysis of test scores is not only a viable alternative, it is evidently more objective, accurate, and cost-effective as well.

*The Education Consumers Consultants Network is an alliance of experienced and credentialed educators dedicated to serving the needs of parents, policymakers, and taxpayers for independent and consumer-friendly consulting. For more information, contact J. E. Stone, Ed.D., at (423) 282-6832, or write: professor@education-consumers.com* produced effects ranging from plus 1% to minus 13%.