

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

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

EDUCATION
next

[home](#)

[unabridged](#)

[past issues](#)

[search](#)

[about](#)

[single issues](#)

[subscriptions](#)

Flunking ETS

by MICHAEL PODGURSKY



How Teaching Matters: Bringing the Classroom Back into Discussions of Teacher Quality

by **Harold Wenglinsky**

Educational Testing Service, 2000

Many education scholars would be surprised to learn that Harold Wenglinsky of the Educational Testing Service (ETS) had discovered, in his own words, “not only that teachers matter most” for student achievement, “but how they most matter.” While many researchers believe that teacher quality is important, in the research literature its influence usually runs a distant second behind the socioeconomic background of students. Nevertheless, based on his findings, Wenglinsky recommended that teachers “be encouraged to convey higher-order thinking skills, conduct hands-on learning activities, and rely primarily upon tests to monitor student progress.” When this Milken Family Foundation–sponsored study was released last October at a highly visible media event at the National Press Club in Washington, D.C., Wenglinsky further recommended that teachers be rewarded for “putting into practice a curriculum oriented toward” these classroom practices, “perhaps through offering advanced certification, such as that of the National Board of Professional Teaching Standards.” Arthur Wise, president of the National Council for the Accreditation of Teacher Education (NCATE), chimed in with a press release of his own: “NCATE is pleased to see empirical validation of its standards.”

Strong claims and policy recommendations should be backed by solid evidence, but the evidence in the ETS study falls short of this

standard. In a nonexperimental situation, where students have not been randomly assigned to treatment and control groups, estimating the impact of teachers on individual student achievement—particularly this or that *style* of teaching—while simultaneously controlling for all of the other household, community, school, and classroom variables that affect student achievement is a very challenging statistical task. This new ETS study illustrates why isolating the effects of teachers is so difficult.

Limitations of NAEP Data

ETS used data from the 1996 National Assessment of Educational Progress (NAEP) in mathematics and science to examine the relationship between students' 8th-grade test scores and their teachers' classroom practices, the professional development they received in support of those practices, and nonpedagogical variables such as teachers' educational levels and whether they majored or minored in the subject they teach. ETS found that teachers' classroom practices, such as using hands-on learning activities and emphasizing higher-order thinking skills, had the largest effects on student achievement, followed by teacher training that supported hands-on learning and higher-order thinking. The only teacher characteristic to improve student achievement was whether teachers majored or minored in the subjects they teach. "These findings," concluded Wenglinsky, "indicate that less attention needs to be paid to attracting certain kinds of people into teaching, and more attention needs to be paid toward improving what our current crop of teachers does in the classroom."

Such a research design, however, suffers from a fundamental "chicken-or-egg" problem. ETS assumes that certain classroom practices cause student achievement to improve or worsen. This is undoubtedly true, but it is equally reasonable to argue that teachers tailor their classroom practices to the academic skills and orientation of their students. For instance, suppose that math teachers with highly motivated students practice hands-on learning techniques and emphasize higher-order thinking, while teachers with students who are weaker or prone to behavior problems focus on rote learning and the basics. A researcher using ETS's methodology would find a statistical association between the teaching of higher-order math skills and student achievement, but at least part of the association is caused by teachers emphasizing higher-order thinking skills with their most able students.

Consider how seriously misleading such an approach would be in a medical study. Doctors choose their treatments based on the condition of their patients. Suppose that doctors treating breast cancer choose

localized surgery for patients with small, localized tumors but radical mastectomies for patients with more-advanced and life-threatening forms of the disease. If we simply compare patients' survival rates with their treatments, without accounting for their previous conditions, we would erroneously conclude that radical mastectomies reduce a patient's chances of survival. By analogy, the bias in a study of classroom practices introduced by teachers tailoring their instructional strategies to their students can be mitigated if measures of students' previous achievement levels are available (with longitudinal data). NAEP, however, being a cross-section survey that only takes a snapshot of student achievement at a certain moment in time, does not provide such data. Without such data, we simply don't know whether emphasizing higher-order thinking skills leads to higher student achievement or whether high student achievement leads teachers to emphasize higher-order skills.

The cross-sectional nature of the NAEP data set creates another problem: the ETS study is unable to account for students' previous school experiences. What ETS has found is that, after (partially) controlling for students' socioeconomic status, achievement in 8th-grade math is associated with several characteristics of 8th-grade teachers. But student math and science achievement in 8th grade reflects not only the contribution of the 8th-grade teacher but also the cumulative contribution of all previous teachers and classroom practices. A positive association between teacher emphasis on higher-order thinking and student math achievement in 8th grade may have as much or more to do with the fact that teachers in grades 1–7 emphasized routine problems and basic skills than with the classroom practices of the 8th-grade teacher. The most accurate way to gauge the contribution, or value added, of the 8th-grade teacher is to adjust for a student's achievement level before entering the 8th-grade classroom. In order to do this, however, the researcher needs to pretest the students or use previous (longitudinal) data on student achievement—information that, once again, is not available in the NAEP.

NAEP's limited data set weakens ETS's findings in other important ways. For instance, the only information on students' socioeconomic status in NAEP is limited to students' responses concerning their parents' educational levels and their home environment (whether a household had more than 25 books and an encyclopedia). This left ETS unable to accurately adjust test scores to account for students' socioeconomic status—a crucial predictor of student achievement. NAEP also lacks data on characteristics of the school, such as size of the school population, grade span, school resources, teacher salaries, curriculum, discipline, and so on. ETS's findings of teacher effects may simply reflect the effects of school- or district-level resources or of policies that were omitted from the analysis.

We simply don't know whether emphasizing higher-order thinking skills leads to higher student achievement or whether high student achievement leads teachers to emphasize higher-order skills.

In fact, ETS omitted from its control variables a singularly powerful predictor of student test scores: race. In the 1999 NAEP data, for example, the test-score gap between white and black students was 0.74 standard deviation. Studies have demonstrated that socioeconomic variables typically explain less than half of white-black test-score gaps. ETS's omission of race from its study potentially biases all of its estimated teacher effects. For example, are those students who are located in math classes with teachers who emphasize higher-order thinking skills disproportionately white? If so, then the positive effects of this pedagogical method may reflect nothing more than the racial composition of the classrooms.

Weak Effects

Putting aside the flaws in the study's design for a moment, let's examine ETS's findings and their interpretation. ETS's data on teachers' classroom practices and professional-development activities are based on teachers' responses to a variety of questions included in NAEP. During the 1996 NAEP, science and math teachers were asked about the types of professional development they had received over the past five years. Math teachers were queried on nine types of professional development. Teachers most commonly received training on cooperative learning (71 percent of teachers), interdisciplinary instruction (50 percent), higher-order thinking skills (47 percent), and portfolio assessment (39 percent). Science teachers were asked about 11 types of professional development. They most commonly received training on cooperative learning (64 percent), interdisciplinary instruction (56 percent), and performance-based assessment (53 percent). Oddly, one of the least common types of professional development for science teachers was using laboratories (27 percent).

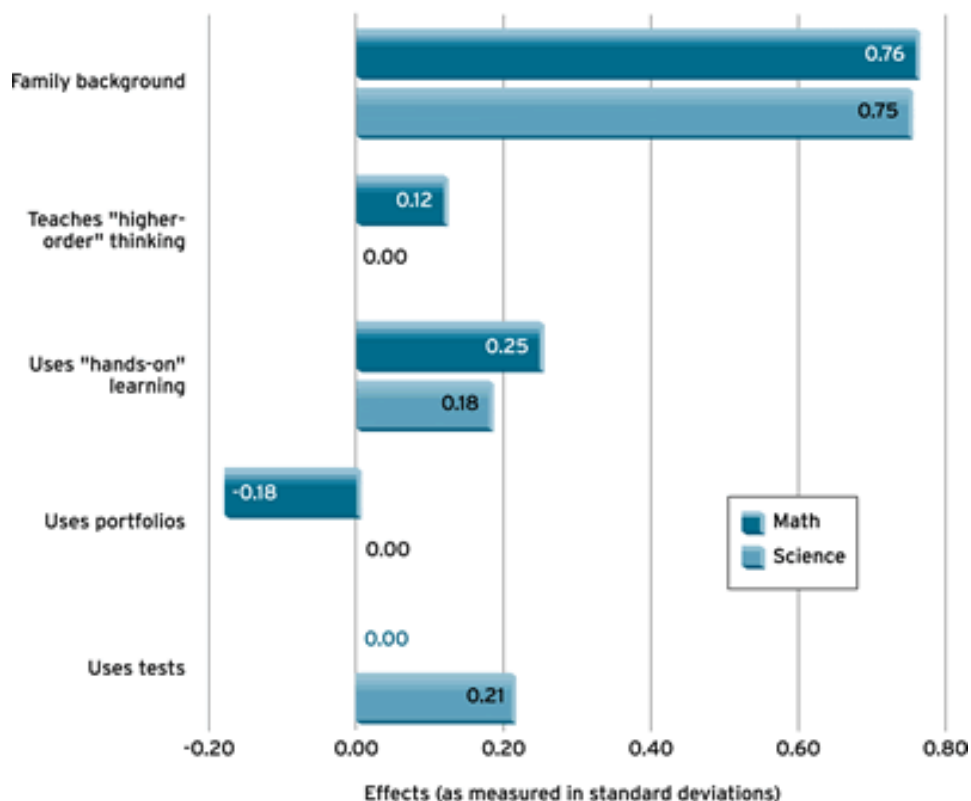
Both science and math teachers were also asked about 21 different teaching practices. For math teachers, examples included whether they "used a textbook once a week" (92 percent said they did), "addressed routine problems" (79 percent), "assessed students from portfolios at least once a month" (18 percent), "addressed unique problems" (52 percent), and whether they had students "write a group paper at least once a week" (66 percent). Of science teachers, 97 percent "addressed concepts in science," 68 percent "worked in groups at least once a week," 67 percent "addressed problem-solving in science," and 18

percent “assess from portfolios at least once a month.” In ETS’s multivariate statistical analysis, it regrouped and renamed these variables. For example, ETS counts math teachers as emphasizing higher-order thinking if they “address unique problems.” Science teachers emphasize higher-order thinking if they “address concepts in science” or “address problem-solving in science.” Several questions about student group activities (cooperative projects, partnering) were combined to measure cooperative learning.

After eliminating many statistically insignificant teacher variables, ETS found several variables, some of which are illustrated in Figure 1, to be significantly associated with student achievement in either science or math. The statistic reported in the figure is the estimated impact of a one standard deviation change in the independent variable on student achievement. Thus, other things being equal, emphasis on higher-order thinking by math teachers raised student achievement by 0.12 standard deviation, but had no effect on science achievement. Of the six areas of professional development, only two had significant, positive effects for math teachers, and only one of eight had a significant, positive effect for science teachers. A measure of total hours of professional development also had no significant effect on student achievement.

Family Background Bests Teaching Practices (Figure 1)

ETS's actual findings show that specific teaching practices have uncertain effects.



SOURCE: Educational Testing Service, 2000.

In general, very few classroom practices were associated with higher student achievement. Of 12 types of classroom practices, only 4 had significant effects for either science or math teachers. One popular pedagogical approach, replacing regular tests with portfolios and projects (“assessment without testing”), was negatively associated with student achievement. As it turns out, most of the professional development and classroom practices popular in schools of education and favored by organizations such as the National Board for Professional Teaching Standards, such as cooperative learning, had no significant relationship to student achievement after controlling for some student socioeconomic characteristics.

On the face of it, these seem to be fairly negative results. The vast majority of investments in professional development bore no demonstrated relationship to student achievement. Even emphasizing higher-order thinking skills is significant only for math teachers. So how can ETS conclude “this study found strong support for the notion that conveying higher-order thinking skills leads to improved performance”? Or what about the conclusion, “This study indicates that the more extended the professional development, the more it encourages effective classroom practice?” The weaknesses of NAEP data are enough to undermine such strong conclusions. But the conclusions are not even supported by ETS’s findings.

Do Teachers Matter Most?

A strongly held belief has emerged in many parts of the educational community that teachers have a very large effect on student achievement—a larger effect, in fact, than that of parents and a student’s home environment. This belief is fueled by the exaggerated claims of the National Commission on Teaching and America’s Future in its 1996 report, *What Matters Most* (i.e., teachers). The statistics on display in Figure 1, however, simply reinforce what many other studies of student achievement have found—namely, that the socioeconomic background of the student has a very large effect on student achievement, an effect that dominates any other measured school or teacher input.

Most of the classroom practices popular in schools of education, such as cooperative learning, had no significant relationship to student achievement.

In student achievement research, effects of less than 0.2 standard deviation are considered small, and effects greater than 0.4 or 0.5 are considered large. By this standard, most of ETS’s teacher effects

would be considered small, and the socioeconomic effect would be considered large. So how can ETS conclude that “teachers matter most”? By adding up all nine of the variables describing teachers and their classroom practices and comparing the total to the socioeconomic effects variable. (In the case of negative signs, ETS reversed the sign. Hence, a good science teacher would need to avoid professional development in classroom management.)

In asserting that teachers matter most, ETS simply means that a teacher who had the right credentials, professional development, and classroom practices could, in theory, offset the effect of one standard deviation of socioeconomic disadvantage. But this is a hypothetical exercise. The study has not demonstrated that such high-powered teachers exist in any significant numbers in the population (or, indeed, that any exist in this sample). Nor has ETS demonstrated that the variation in teacher quality *actually observed* in schools has as large an effect on student achievement as variation in students’ socioeconomic status. That depends on how the nine variables covary with one another. For example, if good practices on one variable are associated with bad practices on another, the two practices might cancel one another out.

In fact, this type of exercise tends to exaggerate the effect of teachers relative to the effects of socioeconomic status. ETS has sifted through dozens of variables concerning teachers’ credentials, training, and classroom practice to find nine variables with the largest effects on student achievement, but has engaged in no similar search for socioeconomic variables. Suppose ETS had a data set with as many good household and socioeconomic variables as measures of teacher characteristics and practices. For example, suppose it had information on family income, size, and composition (e.g., female-headed), data from both parents on educational levels and occupations, and information on the demographics of the communities in which students live as well as the schools they attend. Now suppose ETS chose, from the 30 or so socioeconomic variables, 9 that best predicted student achievement. Without a doubt, the socioeconomic variables would have a much larger combined effect on student achievement than the teacher variables.

Teachers are not randomly assigned to classrooms, and teachers’ classroom practices are not randomly implemented within classrooms. This makes estimating the effects of teachers on students with nonexperimental survey data a daunting task. Researchers have begun to construct longitudinal data files on student achievement by linking student records in some districts and states. These types of projects hold promise for estimating the effects of teachers on student achievement. Of course, experiments with randomized student assignment to treatment and comparison groups would be highly

desirable. Unfortunately, cross-section NAEP data, while valuable in assessing national trends in student achievement, are of limited value in estimating teacher effects on student achievement. Given these statistical challenges, claims that NAEP data can demonstrate that teachers matter most or how they most matter, or can somehow validate education school practices, must be viewed skeptically.

—Michael Podgursky is a professor of economics at the University of Missouri–Columbia.

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

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