

[Education Week's blogs](#) > [Assessing the Assessments](#)

Morality, Validity, and the Design of Instructionally Sensitive Tests

By Invited Contributor Listed Below on June 5, 2014 3:55 AM | [No comments](#)

Today's guest contributor is **David C. Berliner**, Regents' Professor Emeritus of Education at Arizona State University.

Moral Reasons for Using Appropriate Tests to Evaluate Teachers and Schools

The first reason for caring about how sensitive our standardized tests are to instruction is moral. If the tests we use to judge the effects of instruction on student learning are not sensitive to differences in the instructional skills of teachers, then teachers will be seen as less powerful than they might actually be in affecting student achievement. This would not be fair. Thus, *instructionally insensitive* tests give rise to concerns about fairness, a moral issue.

Additionally, we need to be concerned about whether the scores obtained on instructionally insensitive tests are consequential, used, for example, to judge a teacher's performance, with the possibility of the teacher being fired or rewarded. If that is the case, then we move from the moral issue of fairness in trying to assess the contributions of teachers to student achievement, to the psychometric issue of test validity: What inference can we make about teachers, from the scores students get on a typical standardized test?

Validity Reasons for Using Appropriate Tests to Evaluate Teachers and Schools

What does a change in a student's test score over the course of a year actually mean? To whom or to what do we attribute the changes that occur? If the standardized tests we use are not sensitive to instruction by teachers, yet still show growth in achievement over a year, the likely causes of such growth will be attributed to other influences on our nations' students. These would be school factors *other* than teachers--say qualities of the peer group, or the textbook, or the principal's leadership. Or such changes might be attributed to *outside-of-school factors*, such as parental involvement in schooling and homework, income and social class of the neighborhood in which the child lives, and so forth.

Currently, all the evidence we have is that teachers are not particularly powerful sources of influence on aggregate measures of student achievement such as mean scores of classrooms on standardized tests. Certainly teachers do, *occasionally and to some extent*, affect the test scores of everyone in a class ([Pedersen, Faucher, & Eaton, 1978](#); [Barone, 2001](#)). And teachers can make a school or a district look like a great success based on average student test scores ([Casanova, 2010](#); [Kirp, 2013](#)). But exceptions do not negate the rule.

Teachers Account for Only a Little Variance in Students' Test Scores

Teachers are not powerful forces in accounting for the variance we see in the achievement test scores of students in classrooms, grades, schools, districts, states and nations. Teachers, it turns out, affect *individuals* a lot more than they affect *aggregate* test scores, say, the means of classrooms, schools or districts.

A consensus is that outside of school factors account for about 60% of the variance in student test scores, while schools account for about 20% of that variance ([Haertel, 2013](#); [Borman and Dowling, 2012](#); [Coleman et al., 1966](#)). Further, about half of the variance accounted for by schools is attributed to teachers. So, on tests that may be insensitive to instruction, teachers appear to account for about 10% of the variance we see in student achievement test scores ([American Statistical Association, 2014](#)). Thus outside-of-school factors appear 6 times more powerful than teachers in effecting student achievement.

How Instructionally Sensitive Tests Might Help

What would teacher effects on student achievement test scores be were tests designed differently? We don't know because we have no information about the sensitivity of the tests currently used to detect teacher differences in instructional competence. Teachers judged excellent might be able to screen items for instructional sensitivity during test design. That might be helpful. Even better, I think, might be cognitive laboratories, in which teachers judged to be excellent provide instruction to students on curriculum units appropriate for a grade. The test items showing pre-post gains--items empirically found to be sensitive to instruction--could be chosen for the tests, while less sensitive items would be rejected.

Would the percent of variance attributed to teachers be greater if the tests used to judge teachers were more sensitive to instruction? I think so. Would the variance accounted for by teachers be a lot greater? I doubt that. But if the variance accounted for by teachers went up from 10% to 15%, then teacher effects would be estimated to be 50% greater than currently. And that is the estimate of teachers' effects over just one year. Over twelve years, teachers clearly can play an influential role on aggregate data, as well as continuing to be a powerful force on their students, at an individual level. In sum, only with instructionally sensitive tests can we be fair to teachers and make valid inferences about their contributions to student growth.

In this series, see Neal Kingston's blog for more on instructionally sensitive tests [\[link\]](#), and Madhabi Chatterji's blog on validity and test use [\[link\]](#).

David C. Berliner

Arizona State University

Categories: [Validity](#) [High-stakes testing](#) [School accountability](#) [International assessments](#) [Test use](#) [Formative assessment](#)

Tags: [High-stakes testing](#) [International assessments](#) [School accountability](#) [Test use](#) [Validity](#)