

[Education Week's blogs](#) > [Walt Gardner's Reality Check](#)

Using Standardized Tests to Evaluate Teachers

By [Walt Gardner](#) on December 9, 2011 7:52 AM

The demand seems so reasonable: Evaluate teachers on the basis of how much their students have learned. After all, if schools exist to educate, then what's wrong with asking for evidence that they are successful? It's a fair question. The problem is agreeing on what kind of evidence to accept.

At last count, 23 states and the District of Columbia assess teachers in part by their students' standardized tests ("[Nearly Half of States Link Teacher Evaluations to Tests](#)," The Wall Street Journal, Oct. 26). Fourteen more states permit districts to use the data to fire ineffective teachers, according to a report issued by the National Council on Teacher Quality. California is a notable exception. Supt. of Public Instruction Tom Torlakson recently said that the standardized tests currently in use were never intended to evaluate teachers. As a result, any inferences made are invalid.

Although California is out of step with today's trend, it was once in the forefront of evaluating teachers ("[LAUSD faces suit linking teacher ratings to student performance](#)," Los Angeles Times, Nov. 1). In 1971, the Stull Act was passed requiring all state school districts to tie teacher and principal evaluations to student performance. I had been teaching about seven years when I was "stulled," as teachers referred to the process. This required the completion of a form that listed a sample of instructional objectives, ways to measure achievement of these, and plans to grow professionally. Tenured teachers were evaluated by administrators every other year at least twice.

It's important to note that standardized tests were *not* used as the basis for compliance with the Stull Act. Teachers relied on evidence about their effectiveness from a variety of sources, including tests they prepared, student projects etc. The view widely held then was that standardized tests were to be used almost exclusively for diagnostic purposes. As a result, school districts in California ignored the law's most basic provisions. But a new lawsuit filed by a group of parents attempts to force the Los Angeles Unified School District to comply with the Stull Act's basic provision ("[Teachers and test scores](#)," editorial, Nov. 17).

Which side is right?

It's necessary to bear in mind that all tests are not created equal. If test scores are to be used to evaluate teachers, there must be *compelling* evidence to support the validity of their interpretation. Contrary to popular belief, no evidence exists in this regard. In fact, the standardized tests in wide use today are instructionally insensitive, making them inappropriate for evaluating the effectiveness of classroom teachers.

A standardized test can be instructionally insensitive in at least six ways, according to W. James Popham, professor emeritus at the UCLA Graduate School of Education. First, it can be inadequately aligned with the curriculum, leaving even students who were well taught at a distinct disadvantage. Second, it can be too easy, so that poorly taught students still are able to do well. Third, it can be too hard, so that even the best instruction is overlooked. Fourth, it can contain ambiguous items, leading to confusion. Fifth, it can consist of an inordinate number of items that largely measure socioeconomic status rather than inspired instruction. Finally, it can rely overwhelmingly on innate verbal and/or quantitative abilities, thereby measuring what students bring to class instead of what they learned in class.

Standardized tests have their place. They can be used as Finland does to determine if there are systemic weaknesses that should be addressed. They can also be used as one of multiple measures in evaluating teachers. Instead, standardized tests have become the predominant barometer to judge teachers. Tennessee, for example, makes such test scores count half of a teacher's evaluation. This is a mistake, with consequences not yet fully understood by the public.