

## TEST SCORES AND TEACHER COMPETENCY

The Oregonian -- February 4, 2010

"The Stump" Op. Ed. -- by W. James Popham

Teachers should be evaluated according to how well their students learn. This is almost as obvious as saying the winner of a football game should be the team that scores the most points. Indeed, the inherent reasonableness of judging teachers by their students' test scores has spurred many policymakers to demand that students' test performances be the dominant factor by which we evaluate a teacher's competence.

In recent weeks, the push toward test-based teacher evaluation has been ratcheted up remarkably because of the federal Race to the Top program in which states have a better chance of receiving dollars if the state's educational leaders agree to make students' test scores a serious factor in how they evaluate their state's teachers. This is surely not the first time the lure of federal largesse has inclined state officials to adopt a stance that, otherwise, might have been rebuffed.

But judging teachers on the basis of their students' test scores makes sense only if a pair of make-or-break conditions have been satisfied, namely, (1) the presence of clear, teacher-understood testing targets and (2) the use of instructionally sensitive tests. Let's look at both of those necessary requirements, and see why they're so significant.

First, teachers must understand what is going to be tested. It is fundamentally unfair to ask teachers to raise their students' test scores without having a reasonably clear idea of what is eligible to be tested. This would be like asking Olympic gymnasts to perform, but not telling them which factors will be used by the judges who evaluate their performances.

Second, the tests being used must be instructionally sensitive, that is, demonstrably able to distinguish between well-taught students and poorly taught students. Inaccurate estimates of teachers' instructional success will surely be produced if a test can't tell the difference between students who were taught effectively and those students who were taught ineffectively.

If either of these two requirements has not been satisfied, then the use of students' test scores to evaluate teachers is unwarranted. Regrettably, at the moment, in almost all of our 50 states, neither of these requisite conditions has been satisfied. Let's see why.

Currently, most proponents of test-based teacher evaluation want to rely on a state's annual accountability assessments as the tests to be used in this process. The problem with such tests, however, is that they are typically constructed in order to assess students' mastery of a state's officially approved curricular goals.

What's wrong with this seemingly sensible strategy? In a nutshell, most states have regrettably identified far too many curricular aims -- too many to be taught in the available teaching time or to be tested in the available testing time. As a consequence, statewide accountability tests have no alternative but to sample the curricular goals to be measured on a given year's tests. Some curricular goals will be assessed annually;

some won't.

This situation forces teachers to guess regarding which curricular goals will be tested each year. And, of course, a good deal of inaccurate guessing unavoidably takes place. As a result, many teachers end up emphasizing what isn't tested, and failing to emphasize what actually is tested. In most states, teachers really have no clear idea about what's going to be measured on their state's upcoming accountability tests.

If teachers truly understand the nature of the skills and bodies of knowledge being assessed, then they can teach toward such skills and knowledge rather than toward a test's items. Teaching to a test's items is deplorable; teaching to the skills and knowledge measured by a test's items is admirable.

Next, let's look at the instructional sensitivity of the tests that most advocates of test-based teacher evaluation would have us use. An instructionally sensitive test will identify which students have been well taught and which students haven't. But, at the moment, there is no evidence whatsoever that the tests being touted for test-based teacher evaluation are up to that task.

State accountability tests, the annually administered standardized tests used as part of a state's accountability tests, are accompanied by no evidence -- none at all -- that they can tell the difference between students who have been taught well and those who haven't. That's right, there's no documentation that these annual accountability tests are instructionally sensitive. On the contrary, available evidence suggests that today's state accountability tests are instructionally insensitive.

These tests have been constructed using traditional procedures designed to produce comparative score-interpretations, for example, to allow us to say, "Kelly scored at the 78th percentile, that is, outperformed essentially 78 percent of other test-takers." For such tests to provide these sorts of comparative interpretations, however, it is necessary for the tests to produce a considerable amount of spread in students' total test scores.

But to attain such score-spread, many of the items on state accountability tests end up being linked to students' inherited academic aptitudes, such as a child's innate quantitative potential, or to the socioeconomic status of a student's family. Because inherited aptitudes and family status are nicely distributed variables, test items influenced by these factors tend to create the needed spread in students' test scores. Yet, inherited academic aptitudes and family status reflect what students bring to school, not what they are taught once they get there. Many of today's accountability tests are laden with items tending to make them instructionally insensitive.

Can these two problems be addressed so we can carry out defensible test-based teacher evaluation? Absolutely! Serious efforts can be made to communicate upcoming testing targets to teachers. Solid evidence can be collected to indicate whether a test is, in fact, instructionally sensitive.

Test-based teacher evaluation can be made sensible -- but only if we

first let teachers know what's going to be tested, and then make sure the tests we use are suitable for this purpose. Otherwise, with or without federal dollars, test-based teacher evaluation will surely be specious.

*W. James Popham, of Wilsonville, is a professor emeritus at UCLA and past president of the American Educational Research Association.*