

## *Should Test Scores be tied to an Evaluation?*

### **Using state-wide test scores is not an educationally sound solution**

- **Reliable and Valid Data:** The Washington State Institute of Public Policy (an independent research center) conducted a study of Washington's state assessment system and concluded that the strand level data was so unreliable that it shouldn't be used for even large scale building level program decisions. If test data shouldn't be used for program level decisions, it shouldn't be used for individual teacher evaluation.
- **Test Scores Show Patterns Over Time:** While 2-3 or more years of test scores might reliably show student growth or lost educational ground against a particular standard, a single year of scores – which often is suggested as the yardstick in teacher evaluations, is not a reliable measure of teacher effectiveness.
- **Calibration for Growth:** In order for 2-3 years of test scores to be viewed, the assessment itself needs to be calibrated so educators can compare one year to the next. Washington's assessment is not calibrated for this purpose. In other words, the assessment is not designed to show student growth from year to year. Until our state assessment is appropriately calibrated, using state test scores for this purpose is an invalid and a reckless use of data marked by defiant disregard what is educationally sound.
- **Context Matters.** Many factors influence student achievement and, therefore, affect student test scores. These include home support, school attendance, family income level, and parents' level of education. The result is that teachers in wealthier communities are likely to "look" better because their students are likely to score higher on tests. But this is often more a measure of students' home environments than of teachers' instructional effectiveness.
- **Multiple points of Data:** Test scores do not speak for themselves. The judgment must take a variety of factors into consideration. Richard Rothstein's latest book "Teachers, Performance Pay, and Accountability" examines the myth that quantitative measures are widely used for performance evaluation in the private sector and finds that they are not. When educators designed a rigorous evaluation system in Montgomery County, Maryland, they specifically avoided the test scores appearing on a teacher's evaluation. The scores must not speak for themselves because there are too many factors that impact the data other than the teacher, and too much that the teacher does to produce outcomes that are not reflected in the data.
- **Value Added Models:** Though growth based systems may be a more valid way to demonstrate student improvement, recent research has called into question whether these models can sufficiently generate causal estimates of teacher performance. Questions remain about how well they account for uneven student learning trajectories and the nonrandom assignment of students into classrooms. In practice, such measurements fluctuate too much from year to year, require testing in every grade, and exceed the capacity of most districts to carry them out.

## Using state-wide test scores is not a logistically possible solution

- **Limited Tested Grades:** The state level assessment is given in grades 3-8 and 10. How is it logistically possible to connect state assessment scores with educators who teach kindergarten, first grade, second grade, ninth grade, eleventh grade, and twelfth grade? How can state-level assessment results be attached to teachers who have never administered the assessment?
- **Limited Tested Subjects:** The state assessment addresses selected subject areas and does not cover many areas in the school curriculum. Reading, mathematics, science, and writing are the only subject areas assessed on the state-wide tests and even they are not consistently assessed in every grade. For example, students in grade 3 are only tested in reading and math. Science is only tested in 5<sup>th</sup>, 8<sup>th</sup>, and grade 10. How do you assign test scores to educators who teach social studies, health, physical education, visual art, choir, band, orchestra, technology, career and technical education courses, etc? What do you do with librarians, counselors, occupational therapists, physical therapists, certificated school nurses, etc?
- **Over 50% of Teaching Staff are Not Connected to State Assessments:** Combining the grades where the state assessments are not tested and the non-tested subject areas, over 50% of certificated teachers are not even connected with state-wide test scores making a comprehensive system using state test scores for teacher evaluations logistically impossible.

## Possible ramifications

- **Teaching to the Assessment** - If student test scores on a single assessment becomes the basis for teacher evaluation then the test will become the major focus, crowding out broader learning to an even greater extent than it already has.
- **The Campbell Effect:** Generally speaking, the Campbell effect states that when test scores become the goal of the teaching process, they lose their value as indicators of educational status and distorts the educational process in undesirable ways. In other words, the pressure to have students score well on a single test for teacher evaluation becomes so intense that it leads to perverse and unscrupulous practices including:
  - Cheating on the test by both students and teachers
  - Data manipulation (think about the fuzzy math from the state reports)
  - Distorts education by narrowing the curriculum,
  - Distorts education by teaching to the testThe Campbell effect has been demonstrated in public and private sectors demoralizing the workforce charged with carrying out the assessments.
- **Less Collaboration and More Competition:** If teachers are being evaluated based on state test scores, they are less likely to collaborate and help their colleagues.