

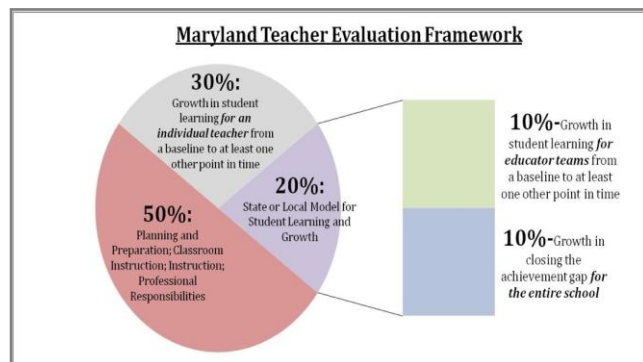
Volume 9, Issue 1, July 2011

**Will Evaluation Lead to Better Teaching?***An Update on Maryland's Teacher Evaluation Process***Executive Summary**

Since August 2010, the Educator Effectiveness Council, a 21 member state-appointed panel, has been meeting to craft an evaluation to determine which teachers are effective in Maryland. The aim is to pilot an evaluation in seven districts—focusing initially on teachers in the tested areas, English and math in 2012. To create the evaluation, the council needed to address two big questions this year: How much should test scores and other data points count toward the evaluation, and what will the impact be of the evaluation results on teachers?

**How Effectiveness is Defined**

The proposal by MSDE for evaluating teachers comes directly from the Race to the Top application which states:



Essentially, the evaluation will have two component parts, one based on test scores and another based on qualitative measures like classroom observations. Within the half of the evaluation that will rely on test scores, one part, or 30%, would be on the change in MSA/HSA scores from year to year, and an additional 20% would rely on a local model for student growth.

However, most districts use state test scores as their local measure for student growth. So, in reality, 50% of a typical teacher's evaluation will be determined by state test scores. There has been a proposal on the council to have test score data over-ride any other data, making it more like 100% of a teacher evaluation.

The argument has been that if a teacher could not show that they could produce student growth that that teacher is ineffective no matter what other work s/he is doing around classroom preparation, outside observations, and commitment to professional development. While there has been no definitive answer on whether test scores would trump all other evaluation data, it is precisely the teacher that is working on his/her teaching but who may not bump up test scores from year to year who needs assistance and support.

While there has been some uneasiness within the council members on relying so much on test scores to determine a teacher's evaluation, the council has voted and agreed to move forward with 50% of the evaluation relying on state tests. The main argument has been that the state must adhere to its commitments made in the Race to the Top application, or put the funding in jeopardy. At the moment the council is working on a matrix that would provide districts a numerical way of calculating the degree to which teachers move their students on state tests. For teachers who teach in non-tested areas, there may be new exams coming so that there will be data on which to evaluate those teachers as well.

**Raising Doubts about Relying Too Much on Test Scores**

Despite Maryland's commitment to using test score data as 50% of a teacher's evaluation, there has been research<sup>1</sup> that has cautioned against placing so much weight on test scores in teacher evaluation. The research indicates:

(1) **Test score data is not always reliable.** There are many factors that go into a student's score on a test, only one of which is the teacher and his/her instruction. It is hard to isolate the teacher as a variable in determining how much s/he influenced the scores. Teachers work differently with different groups of students, and students change from year to

<sup>1</sup> Economic Policy Institute: August, 2010. Problems with the use of student test scores to evaluate teachers; Darling-Hammond, L. December, 2009. "Recognizing and enhancing teacher effectiveness. International Journal of Educational and Psychological Assessment.

year, so it is hard to get a reliable measure of effectiveness. Based on these factors, RAND and the National Research Council have both cautioned that value-added measures using test score data are too unreliable to use.<sup>2</sup>

**(2) Test score data cannot always be collected reliably.** Many districts have problems collecting and maintaining data on student test scores because of the unpredictability of students. Students are absent on test days, leading to incomplete data sets. Also, test scores will be affected by students with chronic absence, lack of access to proper nutrition and healthcare or who suffer from summer learning loss. Conversely, test scores will be affected by private tutors or intensive test preparation.<sup>3</sup> All of these outside factors lessen the reliability of test scores as a measure of teacher effectiveness. The Economic Policy Institute argues that these factors alone should caution any state from using test scores as a measure of teacher effectiveness.<sup>4</sup>

To remedy these problems researchers suggest:

**(1) Not hinging high stakes decisions on the evaluations.** Researchers caution that evaluations should not be used to terminate teachers, or to make the evaluations public. Los Angeles Unified has been the clearest example of what can happen when a district without much forethought makes this data public. Moreover, the pilot evaluation should ensure that districts proceed slowly. Results of the pilot should be used to improve the evaluation tool before any decisions are made about teachers. Moreover, results should be used to improve teachers' instruction.

**(2) Taking multiple measures into account.** Relying on a single measure to evaluate the quality of teaching will not yield accurate data since teaching is so complex.

**(3) Multiple evaluators need to be in place.** Typically, principals' evaluations of teachers are overwhelmingly positive, but they can be very negative in some cases. Balancing those reviews with peer review or student surveys, for example,

would provide a counterweight to principals' views of teachers.

### **What about teachers who teach non-tested areas?**

Teachers of science, social studies, art, physical education, library science, and music teach in what are called non-tested areas. There is no culminating exam by which students are evaluated. How should these teachers be measured then? There are many suggestions for how to do this including peer-review, where peers conduct classroom observations, and student evaluations, in which students are surveyed about their teachers. These methods may also be used for teachers in tested subjects and may be coupled with the use of test score data to provide a fuller picture of how a teacher is performing.

### **What does this research mean for advocates?**

First, advocates will need to monitor the pilot evaluations to ensure that districts are not using the teacher evaluations to make decisions about a teacher's fate and that the evaluations are implemented carefully. Schools could lose strong teachers if an evaluation is not executed well.

Second, advocates should also watch whether teachers are motivated or discouraged to go into teaching because of the new evaluation. Some have argued that relying on test scores so heavily might discourage a teacher to go into schools in low income communities where the struggle to improve academic achievement is so great. At 50%, Maryland is very reliant on test score data, which may or may not be reliable as recent cheating scandals have suggested.<sup>5</sup>

Third, advocates should pay attention to what degree districts are using professional development to support teachers. Baltimore City, Prince George's county, and others have and continue to develop professional development for teachers, but may need to link it to helping teachers meet the criteria of the new evaluation. In the end teacher evaluation should not incentivize teachers to focus more on testing, a bureaucratic hoop to jump through. It should be used to improve teaching.

---

<sup>2</sup> RAND, 2003. *Evaluating value-added models for teacher accountability*; National Research Council, 2010. *Getting value out of value added*.

<sup>3</sup> Koretz, D. 2008. *Measuring Up: What educational testing really tells us*. Harvard University press.

<sup>4</sup> Economic Policy Institute, 2010.

---

<sup>5</sup> Vogell, Heather. July 6, 2011. "Investigation into APS cheating finds unethical behavior across every level." *Atlanta Journal Constitution*.