TEXAS STATE TEACHERS ASSOCIATION / NATIONAL EDUCATION ASSOCIATION
RITA C. HAECKER, PRESIDENT JASON R. HILLMAN, VICE PRESIDENT E.C. WALKER, EXECUTIVE DIRECTOR

TSTA TESTIMONY ON SB 4

While SB 4 has some components that TSTA thinks would be positive improvements, the push to tie teacher evaluations to effectiveness based on the use of standardized test scores is an unsound and unproven approach and because of that TSTA must oppose SB 4.

The bill has some provisions that we believe could have a positive impact on teacher quality. Some of those provisions include:

- Increasing the requirements to become an EC-4 teacher would conceivably increase the quality and effectiveness of such teachers.

- The bill establishes a provisional certificate, which tends to be supported by theory and research that teachers should have a time period in which to prove their effectiveness in the classroom. If teachers cannot show they are effective, then they should no longer continue to teach. TSTA has long supported an induction year, followed by two years of mentoring for new teachers. However, such a program done correctly will not be cheap, but would be worth the money in the long run.

- The bill calls for an evaluation of continuing professional development, which is sorely needed based on what we are hearing from the field. Individualized PD plans are the right way to go in terms of PD. Research is very clear on this issue.

- The review of requirements to be a teacher appraiser is a good idea. Clearly we need to look at the quality and consistency of the training. In fact, we probably need a certificate for this for those who do not complete an education leadership program. And part of the certification process should be performance-based, in which people watch videos of teachers in the classroom and provide written comments and feedback that is reviewed by assessors. This would be costly, but ensure better quality.

- A teacher leader certificate is a good idea, and TSTA supports the development of this concept.

However, the bill has several fundamental weaknesses that cause us to oppose the bill. The major weakness is the reliance on test scores as a measure of teacher effectiveness. Research is pretty clear about this type of effort: it is not ready for prime time yet. There are multiple issues with this suggestion.

However, the major critiques are below:

Weakness 1: Evaluating teacher effectiveness based on test score.

- Most importantly, there is no existing evidence that such an effort will increase student achievement. There is simply no research base for such a policy, regardless of how much "common sense" it makes to evaluate teachers based on test scores.

- Focusing on test scores as the primary measure of effectiveness will increase teaching to the test. As Koretz (2009) notes, teaching to the test results in test score pollution. Test score pollution means that the test scores do not provide an accurate indicator of student learning. Texas already has test score pollution, as evidenced by greater gains on TAKS than on the NAEP. Increasing pressure to focus on test scores will only increase the amount of test score pollution in the system, thus providing policymakers with even less accurate information on student learning.

- The most recent research shows that there is little correlation between teacher effectiveness as measured by student achievement gains and as measured by observations of professional practice. Indeed, the research shows that a teacher can exhibit good professional practice, but factors outside the control of teachers can outweigh the impact of teaching on test score changes.

- It is also important to note there will be no value-added/growth indicator for 2011-12 provided by TEA and perhaps not one in 2012-13 either. So, there will be no data to use that would allow someone to even start looking at teacher effectiveness.

- Will objective criteria be developed for every single high school course? If so, at what cost? If not, then what data will be used to evaluate those teachers in subjects where such an objective measure has not been developed?

Regardless of the strategy employed to determine teacher effectiveness based on student growth, there are serious problems with each strategy that will ultimately lead to lawsuits by individuals losing their jobs based on faulty evaluations. The three main strategies are described below.

Using Changes in Percent Passing

For those choosing not to use a value-added statistical model, most will choose to look at the percentage of students meeting a particular level, such as meeting standard, college-ready, or commended. There are major problems with such a strategy. First, using any proficiency type measure will always provide an inaccurate picture of growth (Koretz, 2009). This is because the distribution of students around the cut score determines growth as measured by the percent

passing. A teacher with a majority of students just below the cut score will likely show great growth even if the students actually don't answer more than a few additional questions correctly. A teacher with all students above passing will show no growth, even if the students answer many more additional questions correctly. A teacher with a lot of students far below passing will show little improvement in the percentage of students passing, even if the students answer many more additional questions correctly. Further, such a strategy does not control for prior student achievement, peer effects, class size, teaching conditions, facilities, student characteristics (poverty, ELL, special education, etc), or school culture. All of these factors have been shown to influence teacher effectiveness.

Using Raw Changes in Scale Scores

Assuming that STAAR has scale scores that are vertically aligned, the alignment would still have to meet several criteria. First, each scale score point must equal the same amount of growth across the entire scale. Thus, moving 10 points at the bottom of the scale must mean the same amount of learning at the top end of the scale. If this is not true, then any "value-added" indicated by scale score changes will be wrong. Also, there must be no ceiling effect. Any ceiling effect—either now or in the future—will destroy the accuracy of this method of determining growth. Further, such a strategy does not control for prior student achievement, peer effects, class size, teaching conditions, facilities, student characteristics (poverty, ELL, special education, etc), or school culture. All of these factors have been shown to influence teacher effectiveness.

Using Value-Added

While using value-added is probably the best option of the three strategies, research has shown the error rates in value-added to be extremely high, especially for beginning teachers and teachers at certain grade levels where there is little historical data. Even under the best-case scenario, 1 out of every 6 teachers will be mis-identified as effective or ineffective. Further, research has shown that the inclusion or exclusion of just one variable can create dramatically different results related to who is effective and ineffective. Most VAM do not have access to all the relevant variables. Moreover, student mobility can wreak havoc on the accuracy of such models as shown by Sean Reardon in his study of Houston ISD. Finally, the test must have the correct psychometric properties (equal intervals mean the same level of learning, no ceiling effect, etc) to make the VAM estimate accurate.

Weakness 2: Time of implementation

There is almost no reliable way to calculate growth across two completely different tests like TAKS and STAAR. Any efforts at calculating effectiveness across two different tests are incredibly complicated and typically fraught with a large degree of inaccuracy. Further, neither TEA nor a district would have ample time to develop a system with appropriate discussion with and buy-in from relevant stakeholders.

Weakness #3: Additional evaluations by principals will be difficult to implement.
Given that many central office support staffers who provide instructional support and mentoring to beginning teachers and assistant principals will lose their jobs, the burden of instructional leadership and teacher evaluation will fall almost entirely on the shoulders of the principal. Yet, principals already report not having an adequate amount of time to provide high-quality assistance. Further, high rates of principal turnover, especially in low-performing schools with high numbers of beginning teachers, will make consistency in evaluations next to impossible. Either the evaluations will be quick and not very accurate or districts will have to somehow find additional money to hire additional instructional support and evaluation staff.

Weakness #4: Changing certification tests requires additional money—money that would be better invested somewhere else at the current time.
Experts would have to be hired to develop the new system. It would have to be tested to ensure validity. Then several hundred thousand teachers would have to be provided training in the new system and thousands of evaluators would have to be provided extensive training in how to deploy. The fiscal note on this bill, particularly to local districts, will be significant.

Weakness #5: The bill provides the Commissioner too much discretion. TEA does not have the knowledge or expertise to craft a teacher evaluation system.
The bill should require a panel of experts to help craft a proposal. The panel should include different types of experts and viewpoints and should have significant representation from classroom teachers.