

Testimony of Kevin Huffman
Tennessee Commissioner of Education

Before the House Committee on Education and Labor

July 27, 2011

Chairman Kline, Ranking Member Miller and committee members, I am Kevin Huffman, Commissioner of Education in Tennessee. Thank you for inviting me to testify about our work to improve education for our nearly 950,000 public school students in the state.

I want to thank the Committee for taking the time to engage in thoughtful discussion about the role that teachers and teacher evaluation can play in the effort to build a better education system. We are grappling with many complicated questions in Tennessee, and I hope that our experiences will be helpful as you consider the broader implications.

Let me start by providing some context about our work. I was appointed by our newly elected governor, Bill Haslam, and have been in this position for a little under four months. Tennessee has been working on a variety of education reforms for much longer, with broad bipartisan and community support. While the current legislature and governor are Republican, the bill creating our teacher evaluation system was passed by a bipartisan legislature and signed by Governor Bredesen, our Democratic predecessor, who did significant work to advance reforms in education. This work has been continued and accelerated by Governor Haslam, who led the effort to implement many reforms, and to pass landmark tenure and charter school legislation this year.

The legislature and Governors have acted in large measure because our education system has not delivered acceptable results. Tennessee ranks around 43rd in the nation in student achievement. At the same time, our state assessments historically showed that around 90 percent of our students were proficient. Additionally, virtually all teachers were automatically tenured after three years, and tenured teachers were evaluated (without data) twice every ten years. The system was broken, and a bipartisan coalition of political leaders stepped in and took action.

Beyond the legislative work, there is broad community support for education reform in Tennessee. While he is known here in Washington for different work, Bill Frist started an organization in Tennessee called SCORE, which pulls together the business, education, philanthropic and local civic organizations under one umbrella to talk about schools. It has been enormously successful in gathering input and building consensus for change in the state.

This coming school year – 2011-12 – Tennessee will launch our new statewide teacher evaluation system. Let me describe how it will work:

- Teachers will receive an evaluation score from 1 to 5, with 5 being the highest.

- 35% of the evaluation will be determined by value-added scores, or comparable growth scores, from standardized tests.
- 15% of the evaluation will be determined by other student achievement metrics, selected through a joint-decision by principals and individual teachers.
- 50% of the evaluation will be a qualitative score based on classroom observation.

These components are in the legislation, and our job at the state department of education is to help districts and schools implement the evaluation system as well as possible.

I want to pause here, though, and note something that I think is important. No evaluation protocol is perfect. There is no system that is 100% objective, 100% aligned and normed, and 100% reliable. One of our great national failings in the discussion about teacher evaluation is that we consistently allow ourselves to be derailed through the lofty and unattainable concept of the perfect system. The reality, of course, is that evaluation in every field is imperfect. The quest is not to create a perfect system. The quest is to create the best possible system, and to continue to reflect on and refine that system over time.

In Tennessee, we think evaluation should be used for several key things. First, support teachers by providing helpful feedback in real time so that they can continue to improve their craft. Second, identify the top performers in the field so that we can study and learn from them, recognize them for their work, and extend their impact by building meaningful career pathways that allow them to touch ever-more kids. Third, identify teachers in need of improvement so that we can tailor professional development to their needs and, in the case of a small percentage who cannot reach a bar of effectiveness, exit them from the profession. Because the national conversation has often focused primarily on evaluation as a means for removal of ineffective teachers, we too often lose sight of the way the vast majority of teachers will experience the evaluation system: as a means for feedback and professional development, and an opportunity to learn from the very best teachers.

As we prepare for full state implementation of our evaluation system this year, we are working on the challenges of both the qualitative and the quantitative components. I will describe briefly how the system works, what the challenges and critiques are, and how we are attempting to address those considerations.

For the qualitative 50%, we field-tested three different observation rubrics and rating systems across the state last school year, with very positive results. We also gathered input from our legislatively appointed TEAC committee – the Teacher Evaluation Advisory Committee – which met more than 20 times over the course of the year to craft policy guidelines and criteria, review field test data, offer ideas about additional implementation needs, and to make recommendations about the quantitative and qualitative data components. This 15-person committee included eight educators, the executive director of the State Board of education, a legislator and several other business and community stakeholders.

Ultimately, we have selected the TAP rubric (the observation tool used in the Teacher Advancement Program) both because of its strong performance in the field test with teachers and principals, but also because TAP was able to provide the level of training and support that we need for the first year of implementation. Here is how this works.

The TAP rubric measures teachers against 19 indicators across 4 domains on a 1 to 5 scale, with clearly defined, observable criteria. Teachers will be observed by principals, assistant principals, or other instructional coaches or leaders designated by the principals. There will be a minimum of four observations a year for professionally licensed teachers, and a minimum of six observations a year for apprentice teachers. At least half of the observations must be unannounced. At least half of the observations must be during the first semester so that teachers get feedback early in the year. The observations vary in length, from full lesson-length observations, to 15-minute walk-throughs, and are followed within a week with both written and verbal feedback.

In order to become an observer, principals and other school leaders must go through rigorous state-facilitated training, and must pass a certification test. We have, this summer, trained nearly 5,000 observers in very intensive four-day sessions led by expert TAP trainers. Each observer then must pass an inter-rater reliability test in which they watch video taped lessons on-line and answer questions to ensure that they understand what constitutes low, medium and high performance on the different components of the rubric. They must also demonstrate the ability to provide high-quality feedback based on the observed lesson by submitting a post-observation conference plan.

On the quantitative side, Tennessee has been collecting longitudinal data on students, with links to teachers, for nearly two decades and has produced value-added scores for teachers in tested subjects and grades for years. For the roughly 45% of our teachers who teach in tested subjects and grade-levels (essentially, third through eighth grade in science, social studies, language arts and math, and high school end of course exams), the student growth component of the evaluation will be based on the same value-added scores that the state has generated and used over time.

For the teachers in non-tested subjects and grade levels, to meet the statutory requirement of 35% of a teacher's evaluation tying to student growth data, in most instances we will use a school-wide growth score for this coming year. For instance, an elementary school art teacher will be rated based on the value-added score of the school for the 35% of the evaluation. Simultaneously, we are working closely with Tennessee educators and technical experts in subject matter committees to identify and develop comparable, alternative growth measures in these non-tested subjects and grades.

Let me identify with transparency some of the critiques of our system and how we are thinking about them.

First, the qualitative observations: In the field test, teachers and principals had an overwhelmingly positive response to the rubric, liked the observation protocol, and in particular liked the forced face-to-face feedback sessions with school leaders. Teachers

felt like the process of observation and real-time, targeted feedback increased their ability to provide their students with effective instruction, and principals learned much more about their teachers' work and how to act as instructional leaders.

That said, there are a number of concerns that teachers, principals and superintendents (generally, ones who did not participate in the field test) have aired in my many visits around the state. First, teachers worry that the observers will not be effective because of skill limitations. We are attempting to address that real concern through rigorous training and through ongoing support. We will have nine coaches across the state who will be going into buildings this year and re-training and helping support administrators who may struggle with the new demands of this system. Additionally, principals are being evaluated this year, and part of the principal evaluation includes an assessment of how well they implement the teacher evaluation. In the end, though, we cannot guarantee that every boss is a good boss. This is true in every profession and every walk of life.

With so many competing demands, principals worry that the time required is too much. The field test demonstrated however, that this should not be a concern. By designating additional administrators and getting them trained through the state program, principals should spend an average of five hours a week observing and conferencing with teachers if they plan their schedules and pace their observations effectively. More importantly, though, this evaluation system propels a critical cultural shift and growing trend in the job description of principals. Principals are no longer simply building and budget managers. They must take responsibility for instruction and for the development of talent in their schools in order for us to meet our ambitious state goals over the coming years.

Finally, the largest challenge I see is trying to ensure consistency in the range of distribution for the observation scores. By this, I mean that we would like the same teacher using the same lesson to get the same score across different schools and across different districts. This also includes achieving a reasonable, consistent relationship between the quantitative and qualitative components for individual teachers across schools, districts and educator groups throughout the state. This level of consistency will not happen without a great deal of ongoing support, guidance and hard work on the part of school leaders, but we are working to build systems and support structures that will allow us to exercise as much quality control as possible.

To this end, we are creating an on-line reporting platform so that principals across the state will be able to enter observation scores in real time, and we will be able to compile data at the school, district and state level. This means that in November, for example, we would be able to see through our state system that the average observation score in County X is a 3.2, while the average observation score in County Y is a 4.2. If the different levels of ratings do not correspond with achievement scores in the district – meaning that if County Y is not significantly outperforming County X on its achievement and value-added scores – we will reasonably assume that the counties are applying difference standards, despite our training and support. We then will be able to engage in site visits, observations, and re-norming of the observers and observation scores. In

essence, we need to make sure to the extent possible that districts across the state are holding themselves to the same bar.

For the quantitative piece, we are proceeding this year with the current system while we field-test and explore additional options for the 2012-13 school year. The biggest current critique is from teachers in the untested subjects and grade levels. Many feel that it is unfair to be assessed through school-wide value-added scores. Here is how we are thinking about that piece.

First, this year we are working with teams of educators and experts to field-test several alternative assessments across multiple fields. For the following school year, we would like to offer districts – at their discretion – the ability to use demonstrated high-quality assessments. Some districts may choose to use these assessments, both because of the assistance in identifying student needs and also for individualizing teacher value. Some districts may continue to believe that school-wide data facilitates team-building and helps create a sense of collective accountability for results.

I will share my own belief on this, which stems in part from my experiences as a former first and second grade teacher. I believe that for academic subjects and grades – for instance, first grade or secondary foreign languages – we should aspire to use assessments that capture teachers' individual impact on student growth. For many subjects, though, - for instance art and music – it is appropriate to use school-wide value-added data. I do not think we should test kids in every single class. Furthermore, teachers who touch large numbers of students in a school have a school-wide impact, not just on reading and math but also on building the school culture that plays a large role in outcomes. As one music teacher shared with me at a roundtable, “When there are budget cuts that eliminate music positions, we are the first people to step up and talk about our school-wide impact.”

An additional concern is that the value-added scores will disadvantage teachers who work in the highest-need schools and classrooms. Our evidence does not support this claim. There are wide disparities in value-added data among districts and schools, and some suburban schools with high absolute achievement scores nonetheless have lower value-added scores. Additionally, as an alumnus of Teach For America, I am proud to note that in our assessment of teacher providers, teachers from Teach For America and Vanderbilt outperformed teachers from every other pathway on value-added scores. Teach For America teachers, of course, teach in the highest need classrooms in the state.

A third complaint involves the volatility of value-added scores. Some experts believe that value-added scores waver too much from year to year. We believe that value-added scores, as used by the state over a period of years, are meaningful indicators of annual progress. To ensure the fairest system, though, we are going to use three-year rolling value-added scores for teachers for their individual assessments where possible. For instance, a teacher who has taught at least three consecutive years will be scored through the average of those years rather than simply through the last year. For teachers with only

two years of scores, we will use the two-year average, and for teachers with one year, that will constitute the score for their assessments.

One additional challenge is that there are a surprising number of one-off situations that impact the ability to use quantitative data. We have teachers who teach multiple subjects across multiple schools, particularly in remote areas, and it becomes ever more difficult to isolate the impact. We have teachers who teach in alternative settings, where students are sent to them because of behavior problems but may only be in their class for a period of a few weeks.

These are real issues, and we care about doing the best job we can in these situations. I feel strongly, however, that we cannot let the outlier examples dictate policy for the vast majority of teachers. We are likely to read many newspaper stories this year in Tennessee that focus on anecdotes about individual teachers who do not fit perfectly within our evaluation framework. We have to strike the right balance of working to improve the evaluation tools for those teachers, while remaining focused on what I believe is a strong system for the vast majority of teachers.

I want to touch quickly on the implication of the evaluation system for teachers. Essentially, what are the stakes?

First, Tennessee's evaluation law states clearly that "evaluations shall be considered in personnel decisions." This simple directive is critical to school district policy moving forward. LIFO – the pernicious system of laying off the youngest teachers first, regardless of how good they are – cannot be used any more. Schools must take the evaluations into consideration.

Second, under Governor Haslam's leadership, Tennessee passed landmark tenure legislation this year. Previously, teachers were granted tenure after three years, and virtually every teacher got it. It was a virtual rubber stamp. Moving forward, teachers are eligible for tenure after a minimum of five years and only if they score a 4 or a 5 on the evaluation for their most recent two years of teaching. Additionally, teachers who gain tenure under the new system will lose their tenure if they are rated a 1 or a 2 for two consecutive years.

I believe this legislation will be groundbreaking for Tennessee over the coming decades. If there is any place for tenure in K to 12 education, it must be tied to teacher effectiveness, not just initially but in an ongoing way.

Let me close with some broad thoughts based on our experience in Tennessee. First, there is no perfect evaluation system. It doesn't exist and we should stop pretending that the goal is perfection. Second, a good evaluation system must have multiple measures. It must have both a tie to quantitative student achievement growth, and it must have multiple means of assessing a teacher, qualitatively. Third, there should be a continuous improvement cycle for the system itself. We are going to review our system every year,

make changes based on feedback from teachers and administrators, and keep making it better.

Additionally, while I have focused on our statewide TAP rubric for observation today, we have approved three alternative observation systems that several districts will use this year. One system is built around ten or more short observations of 5-10 minutes each. Another, through the work of the Gates Foundation in Memphis, uses multiple tools including student surveys. We approved these models precisely because we don't think we have designed a perfect system and because we do think we should have multiple systems in place that we can study and learn from.

Finally, from my experiences to date in Tennessee, I strongly believe that at some point, states simply have to stop planning and dive in to do this work. I know there are many states that continue to kick implementation one year farther down the road. This seems to be rooted in the futile belief that states will perfect the system before rollout, or that opponents of the system will be assuaged by delay. Neither is true. At some point, states and districts have to actually implement the system, and I am enormously proud that Tennessee is implementing the system this year, without giving in to calls for further delay.

Thank you again for the opportunity to present on behalf of my boss, Governor Haslam, and the state department of education of Tennessee. I look forward to fielding questions on this important topic.