

**NC DPI**  
**Teacher Impact on Student Growth**

**Tomberlin: Tom Tomberlin, Director of District Human Resources, NC DPI**

Tomberlin: So my name is Tom Tomberlin. And I am the director of district human resources for the state. And in that capacity I also deal with our teacher growth model, with Standard 6, or EVAAS. And so today I'd like to talk to you a little bit about how we measure educator effectiveness, and particularly how EVAAS fits into the evaluation process along with the observational components.

So this presentation is built on a couple of assumptions that I would like us all to remember as we go through this presentation, and that is that educating students is not an easy task. We realize that. And it's not one that can be boiled down to a simple score or a number based off an assessment. And so we want to keep that in mind as we talk about this, that we are cognizant of that. But we also think that student growth estimates have an important place in educator effectiveness.

It's also built on the fact that we can all improve. All our teachers can improve. And it is through the evaluation process that we are able to give them the information that they need to improve. We tend to shy away a bit from using evaluation for high stakes employment decisions. Though we recognize that has to happen sometimes. But our focus is on how can we use this information to improve the lives of our professionals.

The reason we observe and gather student growth data is to discuss and improve our practice and to improve the learning of students. And it's this point

in particular that we're going to touch on today, how our effective teachers improve the lives of our students, the educational outcomes of our students.

In North Carolina we have a real wonderful opportunity to have multiple years of data now from both our new educator evaluation system and from our student growth or EVAAS system. I was just recently at a convening in Atlanta of the southern states around teacher effectiveness and how they're measuring it. And we are quite a bit ahead of most of the other states in the south region. Most of them were excited because they were piloting a new observational instrument. And I felt like, "Wow, we're in our third year." And we're able to collect data and to compare the data we're getting from our student growth estimates back to observational estimates. And most other states don't even have that data, and they're still dealing with the backlash of measuring teachers using student growth and having a state-wide evaluation system. So I think we're very lucky in North Carolina that we have moved so far down this path and most of the growing pains are past us.

So one of the things when I talk about Standards 6 with teachers and principals and anyone across the state usually we talk about this in relation to the effectiveness of teachers. And I thought this presentation I'm doing is part of a larger message that I'm getting out across the state to teacher and principals and district administrators about how we look at teacher effectiveness. And this time around I thought, "Well, what if we turn this a little bit and talked about what if we thought about effective teaching terms of the impact that it has on students and their learning?"

So what I did was I contacted the folks at SAS, and I said, “Here’s what I want to do: I want to know if there are students across the state who have been assigned to a highly effective teacher three years in a row.” Now, when I say highly effective I mean high value-added. So their Standard 6 was exceeds expectation. And I wanted to know if we had any students that got three consecutive years of teachers with high exceeding expectation. And I also want to know if we have students who got three years of teachers who did not meet expected growth on their Standard 6. All right? And I wanted to see if this has an impact on the student and their educational career.

And the reason I did this is because as I talked to teachers about value-added one of the things that comes up is, “It’s the kids I’m assigned that makes the difference in if I have high value-added or low value-added. It’s the certain grouping of kids in my class.” And so I really thought about this. And I thought, “Well, if that’s the case then the kid is coming through; if the kid is at a certain level it doesn’t matter which teachers they have. It doesn’t matter if they have high value-added teachers or low value-added teachers. They would be the same across multiple years.” And so we wanted to look at this and see what the data says.

Also, teachers talk to me about, “If I’ve got high-performing kids in my class it’s harder to get those kids to grow. So I’m disadvantaged if I have high performing kids.” And on the other side of the coin teachers say, “If I have really low-performing kids they’re not as motivated, there are other life factors that affect their ability to learn and grow. And that’s not fair to me. I’m

disadvantaged in this model.” So what I wanted to do with this analysis was take those kids on the extreme out of the equation. We’re looking at students who scored in the middle of their math or reading tests in the fifth, sixth, or seventh grade. Is everybody with me?

All right. So these are the kids, right in middle of the road kids. They’re not high achievers; they’re not low performers; they’re your middle of the road kids. And we wanted to see if there was an impact for these kids for having successive years of a high value-added teacher or a low value-added teacher. Here are the numbers really quickly. And I have two cohorts here. So one cohort start out in the fifth grade. So fifth, sixth, and seventh grade, 2010-11 to ’12-13. And then I have an older cohort that began sixth grade in that same year, so it’s their sixth, seventh, and eighth-grade years. And I do this because I always worry that if I just draw one set I may be looking at something idiosyncratic, something strange or anomalous. And so if I have two that helps me decide if what I’m seeing is really a trend or just some weirdness in the data.

So here are the numbers: 1,935, about 2,000 kids in the first cohort assigned to either high value-added or low value-added teachers. For the reading it’s about 200, so about a tenth of the size. Now, I don’t know why this is. My guess is because schools across the state ability group [ph] more in math than they do in reading, especially in the middle school grades. But you can see those are the numbers there, but I’ve included them both.

So let’s take a look at what happens to the achievement of these students when they are assigned to these different teachers. As you can see, when we start

out in 2011, the students assigned to high value-added teachers are around the 52<sup>nd</sup> percentile of performance on the test, and the students that are on the low value-added teachers, about 50.5. So virtually no difference between these two groups of students. But every subsequent year the gap between those students gets larger and larger, until the final year, in 2013, we see that the kids who are assigned to high value-added teacher are now around the 59<sup>th</sup> NCE. Everybody know what an NCE is? It's kind of like a percentile. So they've moved from about the 52<sup>nd</sup> percentile to about the 59<sup>th</sup> percentile in three years. That's a large amount of growth as a group. And then you look at the students who were assigned the low value-added teacher, and they've fallen from that 50<sup>th</sup> percentile down to about the 42<sup>nd</sup>.

So what we have here, the situation that we see, is just by assigning kids to different teachers we have created an achievement gap between these two groups of students where none existed before. If you look at reading the same thing holds true. It's a little less dramatic. Our reading scores seem to be a little more compressed than our math scores. But you can see they still start out around the same place. By the end the gap has widened, again, to about 15 NCEs between these two groups on average. And then we have the older cohort, which the group assigned to the low value-added teachers actually was a little ahead on average. But by the end of the year you can see the same trend holds true. And then in reading as well.

As I look at these data, if it truly were the kids in the classes that made the difference in the teacher's value-added score what you should see there is straight

lines going across all three years. Those kids are the same every year. But we see that the teachers that are teaching them have a tremendous impact on their growth over those years. And I would contend that we have created this difference, one, because we don't recognize or we're not using our EVAAS data to help us understand who is generating growth among our students and differentially assigning the kids that need it most to the best teachers. And this is causing this problem that we're seeing up here.

When we think about—I call this an artificial achievement gap because usually when we talk about an achievement gap it's the characteristics that the students bring into the educational system that is driving the achievement gap. But what we see here is we in the educational system have actually foisted this achievement gap on these students.

I'm glad I'm able to present this data because I think it's very important data. But it's really disturbing that this goes on in our districts. And what I would recommend is that we need to take a closer look at when we schedule our students are we sure all our students are getting equal access to high quality and high growth teachers in our districts and not differentially pushing students into either high or low value-added teachers.

So what this tells us is that teaching matters deeply in outcomes for students. We know that this is the only lever we have to increase student performance; it is the teacher. And so it's incumbent upon all of us to be able to recognize who those good teachers are and to provide support and growth

opportunities for those teachers that we find are not meeting expectation. Are there any questions about this before I move on? Okay.

So one of the things that we're trying to do, and one of our basic tenets of the evaluation process is that we strongly feel that the observational components, standards 1 through 5 on the educator effectiveness system, should be providing the information to teacher to grow and develop so that their outcomes on Standard 6 are better. What we don't want to see, and what we strongly urge against, is looking at the Standard 6 rating of the prior year to inform the observational ratings for the coming year. Now, I say that because one of the things we do to see how well that's functioning is we correlate the observational component to the student growth component. So we correlate standards 1 through 5 to Standard 6.

Now, I'm going to just briefly describe what a correlation is. A correlation is a measure of the degree to which two things, two variable, move in relation to one another. So every year I get older I gain one more pound. So as I my age goes up one my weight goes up one. That is a perfect one-to-one correlation, +1 correlation. Some of you may be lucky in that every year you get older you lose a pound. That is a -1 correlation. A perfect negative correlation is one goes up, one goes down. So that's your range here. And right in the middle is zero, which means there's no relationship between the two things of interest. So if every year I got older my weight stayed exactly the same there would be no relationship between the two.

So with that in mind we want to look at what is the relationship between our observational standards and the Standard 6. So the column we want to look at

here—and I know this is hard to see. It's the size of the numbers, not the actual numbers that matter. So in this first column that is the index score from EVAAS. Everybody familiar with what the index score is? If it's two or greater you exceed expected growth. If it's between negative two and two you meet. And if it's below negative two you do not meet expected growth. So that's the range on that. And then we know how the standards are rated, basically a five-point scale.

And so what we say is when Standard 1 moves in a certain direction what do we know about the movement of the index score? And the magnitude of these relationships is about 0.2. Now remember our scale, -1 to 1. And 0.2 is kind of right there in the middle, right next to zero, indicating that there is not a whole lot of information about what's going on with one if you knew the other. So let me put it to you this way. If a teacher came to me and said, "I got distinguished on my Standard 4 rating. Do you want to give a guess as to what my index score for EVAAS was?" I couldn't make a very good guess because that relationship is so loose between the two. I don't know anything about how good your EVAAS score is based off any of your observational components. Everybody with me on that? All right.

So one of the things we want to look at is how are we doing on this and are we getting better. Now, I want to be very clear; the desired outcome is not to have a one-to-one relationship between these two. And the reason is we believe there's much more to teaching than what can be captured in a test score. But we also—if two things are correlated perfectly you don't need both of them; you would just need one of them. If we found that that our observation and our

student growth scores were perfectly correlated we could just do the observations or we could save our principals huge amount of time and just do the tests. But that's not the case, and that's not what we want it to be. What we want is some information about how students are going to perform from our observational component. But we recognize there's a whole lot more going on there that cannot be captured in a test score.

So we want to look at how we're doing. And so now we're very lucky to have two years of data in this regard. And if we look at the '12-13 data we can see that the magnitude of the relationship in that first column there is roughly the same as what we saw in '11-12. It's slightly lower but still on the same order of magnitude; it's still around the same place. The other thing that we need to remember is in the '12-13 year we incorporated about 10,000 more teachers into this model. Because of the North Carolina final exams now we have many more teachers who have both an EVAAS score or a Standard 6 score and a one through five rating.

Now, what I really want to point out, and I point this out to principals quite a bit, is when you look at this first column and see that number, roughly a 0.2 relationship. Then I want you to look at the numbers inside. These numbers here tell you the relationship between any one standard and another standard. So let's imagine that same teacher who came to me and said, "My Standard 4 is accomplished. Can you tell me my index?" "No, I really can't. I have no idea." And they say, "Well, okay. My Standard 4 is accomplished. Can you tell me my Standard 1 rating?" "Oh, yes. I can do that. It's accomplished as well." "But

what about my Standard 2?” “Yes, that’s accomplished as well.” Do you see what I’m getting at here? There is a very strong relationship between how you rate a teacher on one standard and how you rate that teacher on any other of the five standards. It’s so tight that I can probably tell a teacher—if they give me one I can tell them what the other four are, and I’m going to be right most of the time.

Now, what this tells me, and why I think this is critically important is that it tells me we do not measure five distinct standards of teacher performance. We measure one thing five times, and we’re doing it the same way over and over. We’re not measuring discrete practice. We’re measuring this large thing called good teaching; we just measure it over and over and over five different times. Now, this is a problem because it doesn’t help teachers grow and develop.

If you can’t get more granular about what it is about the practice that is not up to standard then it’s really difficult for a teacher to grow and develop. A teacher can be very strong in pedagogy, Standard 4, but be somewhat weak in leadership. That’s possible and probable that that’s going on. And what we need to do is have our principals really dig it into this rubric and finding out the areas where specifically teachers are not meeting the standards as defined in the rubrics. Is everybody clear on that relationship?

Now, I’m not saying that it’s easy. It’s extremely hard to do. And I have actually, in some of the meetings that I’ve had with principals, we have demonstrated to them empirically that there is this thing called the halo effect out there. If I think you’re good at one thing, I think you’re good at a lot of other things. And it is a psychological phenomenon that we’re all subject to. And so it

doesn't mean that they're bad raters; it means this form of bias is creeping in that we have to guard against in our observational process.

And that's what I just said. So there are five discrete standards of teacher practice. But effectively the data indicate that we are effectively measuring one thing five times over and over. As I thought about this as well I thought, "Well, what if we looked at how we rate our teachers in general and are there differences there?" And so what I did was—the problem with five standards the way we have it now, and most of you are aware of this, on the low end we have very little data. 96.5% of our teachers are rated proficient or better. So when you think about, "Well, I wonder what the non-proficient teachers—what does their data look like?" well, that's really hard to say because there are so few of them across the state.

So what I did to do this analysis is I lumped all the teachers who have a developing or not demonstrated anywhere on their evaluation, I called them less than proficient. So they're this first group right here. And then if a teacher had no developing or not demonstrated but at least one proficient score I put them in the proficient category. And then if all the ratings were accomplished or better they were in the final group to the right. The blue is what we looked like in 2011 and '12. And the green is '12-13.

So what I did was say, "Given those categories, what does the student growth look like for those teachers as a group?" This first group, their average index score is -1.44, which we would think is about right for that group. We're saying that there is some deficiency in their performance, and we would expect

that their student growth would be somewhat lower. Then we look at the proficient group, a -0.28. So this group, again, right at the zero mark but just a little smidge below zero. And we think that's about where they need to be. And then the accomplished or better, 0.626. So on average they are performing better than the proficient group. And that little asterisk beside it tells me that I know that difference is meaningful. That's not just a difference in the number. It's that difference is statistically significant.

You can look at the '12-13 data and see that the trend continues. So in the aggregate, on average the higher we rate a teacher the higher their student outcomes are. And that's a very good thing. That's a wonderful thing. Averages are really good for federal reporting and for state presentations. For schools they're not so good. And here's the reason why. This is where the average is. That little line up there next to the number. But these are actually teachers. These little dots down below those boxes, those are teachers in those categories. And that's their value-added score. So really large negative index scores for teachers in those categories. And so as we think about—as we simultaneously kind of congratulate ourselves that we are able to show that teachers with higher ratings, on average, have better student outcomes and it's a great thing.

This is where the work lies for us. What is going on with these teachers? How can we help generate the student outcomes that are consistent with the evaluation ratings they're receiving. Now, I'm going to be very clear, this is not about rating teachers differently. I'm not suggesting that those teachers in the proficient category who are below that line, that we need to change their rating.

But what we do need is to change what we're doing in the evaluation process for them. Are we giving them the guidance they need to grow and develop and achieve the outcomes that I would argue they want to get for their kids. I don't think these teachers are like okay with having this large negative growth.

The other thing I'm going to look at here is in this accomplished bin. This teacher, to me, is very remarkable, is in the accomplished or better, and has the largest negative growth estimate in the state but rated accomplished or better on all five standards. Now, my thought when I saw that was not, "Oh, yikes we got that one wrong," or, "We need to do something to get that teacher out of there." My thought was first, "That's me." So I'm going to make it me. This guy teaching high school Latin—this is what I taught when I was in high school. We're telling him that he is great, he is fantastic, there's nothing I can see, there's nothing that you need to do differently for your kids. But he's achieving outcomes that are very large and negative. How does he reconcile that information? He sees that EVAAS score. He logged in. He knows what the deal is. What are we doing as the district to support his growth so that can look at his student outcomes and feel the same pride in his teaching and knowing that he's delivering for his kids that these other teachers in that bucket who are up on the higher end probably feel.

So these graphs, somewhat difficult to look at, but one of the things we know is that standards 3 and 4 correlate most strongly with the EVAAS score than any of the other five standards. So I wanted to point out again, as we go across the rating categories for Standards 3 and for Standards 4 we see that same general

trend. The higher we rate our teachers, on average, the better their student outcomes, their student growth estimates. But even so we could see the significant areas of work for those teachers on the lower part of that what's called a box-and-whisker plot. Those teachers that are extending down, that's where we need to dig in and find out what's going on, what can we do to help and support those teachers and continue to help them grow.

Now, the final thing that I want to present here is—and this came to me after reading a research article where in Cincinnati public schools—and basically they had an evaluation system that was binary; it was two categories: you're horrible or you're fantastic. And so, as you can imagine, 99% of their teachers ended up in the fantastic category. And very few were in the "you're horrible" category. But there wasn't a lot of—it was very much a checklist; there was not a lot of feedback and not a lot of conversation, and they switched to a new rubric that had much more expanded areas of practice. And a much more robust rubric, and part of it was to have conversations, have post summative conversations with the teacher about how to improve practice.

So this sounds very familiar to me. When I read the article I said, "This is North Carolina three years ago." We had just finished using the TPAI, which was very much a checklist. And we had switched to this new rubric that has much more depth to it and relies more heavily on feedback on performance. And what the authors of the research study had found was that when teachers switched from this checklist, no feedback evaluation system to the new evaluation system their

student growth started climbing. And it consistently climbed for about six years after the introduction of that evaluation system.

So I thought, “What if we looked at this from our perspective?” Now, I can’t do that whole big study. I don’t have the access, and quite frankly I don’t have the smarts to do it. But I was able to go back and say let’s look at those teachers in those big three categories again. How we rated them in 2011-12, when we told them that they were less than proficient. What happened to their growth from the ’11-12 school year to the ’12-13 school year? Does everybody understand what I’m saying?

I just looked at those teachers in their evaluation container; I put them in those three containers of evaluation and said, “On average, what was the difference in their growth from ’11-12 to ’12-13?” So the teachers who were classified as less than proficient, on average they grew four-tenths of an index point. That is enormous. That is a huge amount of growth. On average that’s how much they achieved. Then if you look at the proficient, when we told them they were proficient in ’11-12—and teachers don’t like to be told they’re proficient. They think that’s a C, that that’s an insult. When you told them they were proficient their Standard 6 rating, their EVAAS score moved on average two-tenths of an index score.

Now, I want you to look at the size of this group too. Almost 10,000 teachers, 9,500 teachers in there. And on average they moved up that much. Now look at the accomplished or better group, the largest group in the ’11-12 school year. When we told our teachers, “You’re great; you’re fantastic; don’t

change a thing,” what did they do? They did not change a thing. Their average EVAAS growth is about zero.

So I would submit to you that it’s not the rating that these folks got. It’s not whether they got accomplished or whether they got developing. It’s the information that their principals gave them: “I need you to do this better,” specific feedback on performance that generated this growth. And I would contend even if we’re rating our teachers as distinguished we can be telling them something to do the next year to increase performance.

It’s no secret why—everybody has seen the teacher growth curve? Yes? We’ve all seen this? First five years, tremendous growth, and then after the fifth year teachers on average stay exactly at the same level of performance for the rest of their career. And I would contend that it’s because we quit telling them, we quick pushing them, to do better. There is no growth in nature without adversity, without something to fight and grow against. And if you’re not giving that to even your highest evaluated teachers they will not grow and continue to grow their students.

So, some rhetorical questions that I’d like you to ponder before we move on is: how good of a job are we doing at recognizing the practices that lead to better student outcomes in our schools? And I would say on average we’re doing okay, but we all—I would hope that I’ve given you some ideas about how to go back and look at your own data in your own districts to say, “So where’s the work? Where’s the work that I can do?” And you have these data. Every district has gotten teacher-level EVAAS scores now for the past three year, and you have

your evaluation scores. You can do these kinds of analyses to look and see where are the discrepancies and can I find out how to help these teachers grown and improve.

There's the conventional wisdom that critical feedback for teachers is demoralizing. I often talk to principals who say, "I know I should be telling them something, but I'm afraid they'll just quit. I'm afraid they'll curl up in a ball and they won't teach and I'll never get anything out of them." And we're finding the data suggest that that is really not the case, that when teachers hear this critical feedback they may not be happy in the moment, but the evidence suggests that they respond to this in very positive ways, in ways that are extremely positive for their students.

That's the end of that data presentation. This is some logistical stuff about where we're moving with educator effectiveness. But I wanted to see if there were any questions about any of the data that I presented in this section. No? All right.

[END RECORDING]