

The Teacher To Teacher Initiative was created by the US Department of Education to provide the latest strategies and research on educational practices that work inside a classroom. School really matters and it matters for everybody now. Students need to understand where they stand relative to the standard so they know where they have to go. Are you teaching or sorting. If you are grading on a curve, you're sorting. And what message does that send. This series features teachers from across the country presenting techniques that can be used with students of all ages. It's just one way the Department of Education is helping teachers get the support they need so that no child is left behind.

(MUSIC)

My name is Peter Robertson and I'm from Cleveland, Ohio, where for the last 5 years I've been in charge of technology, assessment, evaluation and accountability. I was asked to build the standards-based report card in Cleveland, and much of what I'm going to show today comes from the lessons we've learned over the last 4 years – 3 ½ years – in doing that. But the piece of my introduction that's more important to me is that I was a high school history and government teacher in Montgomery County, Maryland, and even before I really understood what standards were, I think I understood the value of standards-based education because I taught to two tests there. One of them was the Maryland 10th grade citizenship test, which was a graduation requirement. Kids had to keep taking until they passed it, or they didn't graduate. And it was a terrible test. It asked questions – you had to know things like Bill of Attainder and writ of habeas corpus. With a school that was ¼ ESL, I had students who would take that test repeatedly and fail it repeatedly, even though it was clear to me that they were going to be great citizens. So, I turned my course into drill-and-skill against my better judgment. I also had the privilege of teaching to the AP US History test – and I say privilege because teaching to that test required me to be the very best teacher I could be, and required my students to be the best historians they could be. And I truly believe that you can write tests like that at any grade level in any subject area. And when you have tests like that, teaching to the test is a privilege, I think. So, I'm a big advocate of outcomes-focused education. And the standards-based education efforts in this country are enormously important. We're not very far along in the process, and in most instances we tend to overestimate change in the short run and underestimate it in the long run. So, we've got lots of frustrations with what's going on. But, I truly believe that it's an essential and seismic shift in public education and it has implications in the way we communicate to students and to parents and the way we grade students. And that's really what we're going to talk about today. I have four objectives. They are first to explore the implications of standards for student report card and teacher grading practices, because I think there's a lot in our practice that we need to think hard about.

Secondly, discuss some practical ways to integrate standards-based instruction with student involved classroom assessment grading and reporting practices. And I don't pretend that we'll get very far on that today, frankly. This is not professional development, right? This is a session where I'm, I hope introducing you to some ideas, provoking your thinking, providing you with some guidance to some research that you can take back for discussions in your buildings, and in your districts, which is, of course, the real professional development – not listening to me, and engaging with me, and stuff like this. I do want to review one strategy for using grading practices to keep the focus on standards-based education in an era of high-stakes testing, and bottom-line is because I know this isn't professional development I want to encourage you to put together study groups to explore effective student-led classroom-based assessments and their implications for grading and reporting back in your districts. What I propose to do today is spend some time on why standards-based grading matters, spend a little bit of time on some standards-based don'ts, and actually do a few exercises to get at that, talk about overall how you go about building a standards-based grading system, and I strongly believe that whether your district, or school, are involved in this, there are things that you, as an individual classroom teacher, can do, kind of regardless of what external policies and procedures are being imposed on you. I think it's really important to touch on this issue of progress versus achievement, because that's really the central conundrum of grading and reporting. Students need to understand where they stand relative to the standards, so they know where they have to go. But, by the same token, you need them to feel good about progress so that they're motivated to continue. Finally, I just think that in this day and age it's important to keep anchored in the testing world, and I want to talk a little bit about that. So, first, on this question of why standards-based grading matters, I apologize – I have a little bit of a stump speech I have to give, and it's me and it's why I think this topic is so important. This is a quote from the law – No Child Left Behind – stating its purpose – “It is to insure that all children have a fair, equal, and significant opportunity to obtain a high-quality education, and reach, at a minimum, proficiency on challenging state academic achievement standards and state academic assessments.” Now, leaving aside the question of implementation, and even the question of whether you've got a challenging state academic assessments that are well-aligned to high quality standards – and I recognize those are all big issues – I hope you agree with me that this purpose is absolutely right. And, in fact, it's because I think we are so far from having the right assessments, that the role of the teachers' efforts to assess students is so critical. You know that AP US History Test doesn't exist for all of us in all of our subject areas at every level. And so in many ways, you have to be the one who gives students those opportunities in ways that are more meaningful than those state tests. But this

overall purpose really reinforces something important about our schools. Our schools were invented 100 years ago or so, based on some pre-existing models, to help us cope with industrialization and immigration, and urbanization. And, their purpose was fairly well-suited for the time, but frankly their purpose was to sort students, and to socialize those who were not pulled out to go on to college to be prepared for fairly scrutinized work in the factory. I mean the bells are as much of the educational process for the factory worker as some of the stuff that is going on subject wise. So 100 years ago or so only 10% of Americans graduated from high school. So sorting was absolutely the goal. Now, today those factory jobs require statistical process control, and all kinds of other critical thinking skills and knowledge base, so we now need – and frankly just to even survive in our society with its complex finance decisions that you have to make in your personal life and web access and all the other things you have to do to be a fully functioning citizen – we need a level of skills and knowledge and thinking ability that for every citizen that we didn't used to need. So even though our schools have come a long way toward educating all students, we have a long way to go. In the fifties, ½ of Americans graduated from high school. We like to talk about “the good old days”. Well, the good old days never existed. It was easy to teach a classroom of 45 kids when you kicked out the half who didn't fit. Today 85% of Americans graduate from high school. That's much better – there's been enormous improvement in public education, and you all deserve some of the credit. Unfortunately, the world economy is moving faster than that and we need 100% to graduate, and we need them all to graduate with really high-level skills. So, we've moved from sorting students to educating all students. And that's changed the emphasis in the schools from what's taught, to what's learned. Again, if you're sorting you teach, and whoever seems to pick it up is tagged as a success and moves on. Well, today that's not good enough – it never was good enough, but it worked for the economy. Today, all of our students need to learn. And finally, school success doesn't predict lifelong earnings. That used to be true. One of the co-editors of the book “The Black-White Test Score Gap” had done some research back, I think, in the late sixties, looking at Air Force Vocational Aptitude Tests achievement batteries, and looking at long-term income of people. He drew the conclusion that those test scores in no way predicted lifelong earnings. But, when he edited the black-white test score gap he said current evidence suggests that that's no longer true. Test scores and any other indicators we have of student achievement now, in fact, do predict lifelong earnings. And we know that, from our communities, the high school graduate's income has been stagnating in this country for the last 30 years. High school drop-outs' income has declined. It's only college graduates who have improved their average economic – their earnings. And that's not because they're college graduates, but because college signifies a set of skills that we're getting

better at measuring. So school really matters, and it matters for everybody now. Now it matters so much that I would argue that education is the civil rights issue of the 21st Century – and this is my soapbox piece. And just as voting rights was the linchpin of the civil rights issues of the past century, I would argue standards are the linchpin of the civil rights issue of the 21st century. And I say that because implicit standards are where you get inequality. If standards are vague, if they're not clear, if everyone's not clear on what they need and how to get there, inequity breeds itself. When we know that – again, going back to the last civil rights agenda – because literacy tests were used to deny voting rights, the typical mechanism was you had to prove you could read in order to vote. If you were a poor, illiterate white farmer, you could pretty much count on the constitutional section that you had to know in order to read, so you could memorize it in advance, you'd show up at the courthouse, you'd be asked to read, you'd look at some scribble on the page, but you'd have memorized that line, you'd read it off, and you'd get your voting card. If you were an African-American farmer and you had the nerve to show up at that courthouse, you would have a much denser piece of the constitution, you wouldn't have any idea in advance what it was, you'd be tripped up on whatever you said, and denied the right to vote. Similarly, anytime we talk about hiring bias – are women being hired in large enough numbers to be firefighters? Well, when that happens there's almost always a vague job requirement. Because if you define the job requirements clearly, that's not possible. If you turn it from “women can't do this job” to “you have to be able to climb a ladder with 115 lbs. over your shoulder in x number of seconds” there's no question – there's the standard. Lending inequity – you know, good ol' boy lending practices – are where we get in trouble. I like the guy! Well, that's not only bad for equity; it's also bad for business. People who have clear scoring practices for loans do better business, and they also don't get slapped with red-lining suits. And, of course, for us educational access blocked by admission grading and promotion standards that aren't clear. So, I'm going to argue that standards are the linchpin of the civil rights issue of the 21st century, and quality assessment makes those standards clear and provides data on what's been learned. Now, there are lots of gurus that go around the country talking about data-driven decision making. But they all basically make the same point. If we've got academic standards and we want them to influence instruction, and we want to make sure instruction brings children up to those standards, then we've got to have assessment. And we have to look at the results of that assessment to understand what we know. We have to communicate the results of that assessment to students, to parents, to other colleagues, and this is kind of a constant process and if we do it well we're going to have better instruction that brings us to standards. Now, the point I want to make about the importance of report cards is that there are

really two ways of doing this process of assessment analysis communication. You can do it in a formative way that provides descriptive feedback to kids. The goal here is not a score – the goal is to understand what the student knows, to help the student understand what the student knows, to help the student understand where they need to go, and give them some pointers to get there. So it's a conversation, and frankly you'd like them to have questions that they get wrong so you can learn from the process with them. You want them to do a first draft that's not very good so you can work it through with them and help them understand what they need to do better. And you know that – I mean, that's the core of your teaching, and we know there's lots of research that says good formative assessment does more to improve student performance than almost anything else. Think of it as helping a kid through the woods. If you've got a clearly marked path and along the way you're offering them tips like “stay out of the quick-sand” and “walk over here” and “you've got to speed up a little bit” they're going to feel more comfortable, they're going to feel more confident, they're going to work harder at getting there. Summative assessment – grading and reporting is something we have to do. At some point we have to say this child is accomplishing what needs to be accomplished and therefore goes to the next level. This child needs help that I'm not able to provide and whether that's a series of interventions, right up to an IEP, it depends on the situation and condition. But, you definitely need, at some point, not only to communicate expectations and how well or poorly the child's doing on those expectations, but also, frankly, to help the school district make resource allocation decisions for that child. So, the critical issue in grading is that when you do report cards, or whatever you call them – the reports home – you are the one straddling these two worlds. You've been doing the descriptive feedback, and now you have to say some of the things about that child. And so that process – it doesn't happen anywhere else except in the classroom with the teacher, and I would argue the focus therefore ought to be on what is grading and reporting, how do I do it, how is it different from other things I do with students to provide them feedback and assessments. So, standards-based grading “don'ts”. I'm just going to lay out a couple of things for arguments' sake. And then we're going to do a few activities to kind of play with what that means. And my basic argument is if you think about standards-based education, and you really explore what it means, and you think about what I just said, there are a couple of practices that we're involved in that really don't make any sense. First of all, grading on a curve. Does anyone in this room work in a district that has a policy that says you must grade on a curve? That's a relief. Most districts blew that out in the 70's, and someone told me that Arkansas had a state law that says you cannot grade on a curve. It also says you cannot consider behavioral factors in your grades, which is my 4th point, which is great. Of course, the law and practice are two very different things. Raise

your hand if you live in Arkansas and were not aware of that law. I wouldn't be surprised if any were. Grading on a curve – are you teaching or sorting. If you're grading on a curve, you're sorting. And what message does that send? If the standards are clear, you know how to assess against those standards and I reckon that's a big issue we'll come back to. Then if 100% of your students are meeting or achieving that standard, then 100% of your students deserve whatever grade or mark goes along with that. And if none of them do, then none of them do. But to look at that group of students that frankly randomly assembled in your classroom this year and it's different from last year and say half of them are going to get an A – well that's completely antithetical for standards-based education. Secondly, implied or emerging expectations. My favorite question there is when did you first draft the final exam?

And I was, again, a high school history and government teacher and I've got to confess; there were a couple of years when I wrote it the night before. It's okay to be editing it the night before – it's not okay to be writing it fresh the night before. Frankly, you ought to write it before you start the course if you give a final exam. If you give some sort of summit of assessment, you ought to have an outline or draft of what it is before you start teaching, because if you don't know exactly where you think you're going in terms of what students ought to be able to do, then you can't communicate to them what the expectations are. And then, you know what – you're doing all the work, and they're the ones who are supposed to be the workers. But in order to be a worker, you've got to know what you're supposed to accomplish. Third, all assessments tasks and/or scores are equal. If you believe in student growth, that doesn't make sense. It ought to be what have they accomplished, not where did they start – and we'll do an exercise in a second that gets to that. Fourth point – fourth don't – including non-achievement factors. And again, I'm again told that Arkansas has a state law that says you can't do that, and I think that's right. If Peter is a fabulous geometry student – he can solve every proof, he understands all of the concepts and definitions, he's demonstrated achievement against every one of the standards, and Peter is a little jerk. Peter still deserves the grade that goes with achievement. Now, I might – as his teacher – if I don't have a place to put behavior scores on that report card, write a note home – maybe even try to get a note in his file that says Peter is terrific at geometry, but I doubt he's employable – but that's not what I'm being assessed on. In fact, actually, some school districts have employability grades. They actually take all those behavioral scores from different teachers and they compute an employability grade at the high school level – it makes a little less sense at the elementary level. But, because that's a valid thing to report on – attendance, effort, responsibility, etc. – but it's not academic. The only issue in academic standards is can Peter demonstrate the knowledge skills, etc. that the standards are asking. And finally I

would argue a standards-based grading don't is the composite average score. How valid is Peter got an 83%? Two geometry teachers – one is teaching basically 6th grade level math, giving basically 6th grade level assessments, everybody's getting 100%. Another teacher is introducing geometry and the kids are struggling to get 40%. You know, 40% and 100% aren't comparable there, which brings us back to the issue of standards. So, enough me talking, let's do a little exercise to get at this. You all are teachers now in a parachute packing course. And you have three students and you decided the best way to assess them was to actually have them pack a parachute 7 times throughout the course. And each time you assess them on a checklist of parachute packing attributes, and you know from experience that if you get a 60 or higher on that checklist that parachute is guaranteed to open when it's supposed to. So mastery is a 60%, but the checklist can go as high as 100. So you have 3 students in this class – student A, student B and student C. And here are their scores on those 7 parachute packing assessments. For the numerical of you, you can look at it. For the visuals, here is student A in red – here's student B in blue – here's student C in green – and here's the mastery line. So my question to you is which student would you use to pack your parachute? If these were scores in the typical teacher's grade book, which students would pass – which students would fail? Is there any discrepancy between your answers to questions #1 and #2 – if so why does this occur, and finally what do you think the implications are of this are for the way you calculate student grades? So going back to this, talk to your neighbor, think pair share – we'll have a conversation in a minute. By the way, don't get out the calculators, I will tell you now all of these add up to an average of 70.

“This one would be going – have an A one time and a C the next time and a D and then a B, then a C – back and forth – where the other one is gradually going from, if you consider grading, a 40% up to a 90%.” “If you report after 9 weeks what you should be reporting is that performance at the end where naught average”

Okay, let's go ahead and do a poll here – raise your hand if student A is packing your parachute – raise your hand if student B is packing your parachute – raise your hand if student C is packing your parachute. Okay, anyone want to tell me what you all discussed about why A wouldn't pack your parachute? Unreliable. You know, it's interesting. Any idea why that might be? Anyone here consider the possibility that this is your fault? And you know another possibility that someone brought up – you do this session enough times you get all kinds of interesting comments. One possibility that someone brought up is actually this student gets it, knows they get it, and is experimenting with different approaches, because their understanding is that education is an opportunity to experiment and have a chance to succeed or fail, and if you had told that student in that case that this is your parachute tomorrow they would have gotten it right, but they're just

experimenting with different techniques. So, you know – which student do you most want to claim is yours? B is your student, right? Growth. She didn't understand and now she does and I taught her, right? And I brought up this issue of whose fault it is not because we can know that stuff obviously from a piece of paper like this, but because one of the things that standards does is it forces us to ask is what we're doing working? If not, why not. It's really hard. All helping professions – I'm including doctors in this too – your day is full of helping your client, helping your student, and you know you are doing the very best you can to help your student. And when you go home you're exhausted, and you have all kinds of evidence that you've been helpful. So whether you've been as helpful as you possibly could be is not something that naturally enters your head the same way it does if you're working on a manufacturing line and watching products go by and some of it's good and some of it's bad it's much easier to think gee, why is this one not as good as this one? But that's not what you do. You're engaged in working with clients, so it's very hard to deal with this question of is this my fault or not? Is there something I'm doing wrong, or not? And I don't mean to put it in fault and blame, but you know standards-based education requires us to ask why isn't every student learning and what could we try differently, how can we work with students to figure that out. What do you think is going on with student C? He memorized it and forgot it, or bored. Now, I want to come to this issue of exiting him because this is an important – I'm here to talk about how we'd like to change things. There are limits to what we can change. And one of them – standards-based education turns the whole notion of time is fixed, results vary on its head and says here are the results we have to achieve, time must vary. And that, by the way, includes if a student can do what they're supposed to be able to do before the course starts, they shouldn't have to sit through the course. But we've got this history, certainly at the secondary level, of Carnegie units and timelines, at seat minutes. It's completely antithetical to standards-based education and we're not going to change that part unfortunately. That's why differentiating instruction is so important, because you've got kids who frankly should be exited right out who you're going to make sit through something. So what can you do to help them find another way to grow in that time period? My favorite example is there's a high school that spent years taking standards seriously and developing a set of outcomes and it was exit by performance. Some of them were writing tasks, some of them were speeches, some of them were projects, and they had an exit team that interviewed them, and when they were done with this list of 40 or 50 that students had to do and the scoring rubrics and all of that and they could hand the book to a kid and say here's what you've got to do to get out of here with this diploma, then they were able to say – hey, why do we have courses in this place? Let's scrap all of that and say, "How do we reorganize adult time to help students get there?"

And in some cases there was still a geometry class, because a group of kids all needed the same stuff, but in other cases there was a lot more free time, there was a lot more of students working with teachers on a specific set of issues – let’s get this group together on presentation skills to prepare for their defense. It was a great system and this was an affluent school district and a new school board threw the whole thing out, because despite dozens of letters from competitive colleges saying they loved it, the parents in this community believed their kids were going to be at a disadvantage for getting into college. And fundamentally this absolutely correct way to think about the implications of standards based education just couldn’t get through the culture. So there are limits to what we can change, and students see as an example of it. But I really want to focus on the grading – what grades did you give these kids? First of all, in a traditional teacher’s grade book what grades would these kids be likely to get? What would you give the blue one? Raise your hand if you’d give student B and A. I want to explore that question a little more, because we all recognize who we want to pack our parachute, and we all recognize what would typically happen, and I think we understand the discrepancies. But, what you do about that is difficult, because we all know about averaging. And if you’re going to average grades, everybody understands it. No matter how wrong we now understand it to be, it’s a lot easier to defend that in a world where everyone’s been averaging forever, than it is to do something else unless you’ve got a clear set of rules. And what I’m going to argue is that your students, your parents, don’t need traditions if you can communicate clearly up front that you have rules for this. And by the way, this is different for elementary and secondary. Elementary folks I recognize are more likely to take you on faith – to need less from you in terms of here are the rules I follow to assign this grade – because frankly the stakes are lower, or are perceived to be lower. But I would argue that the same care and concern ought to be given to having rules for how you calculate grades at any grade level, but they don’t have to be the traditional rules. If you’ve got good rules they can be whatever you want them to be. And that’s what I want to explore with this next task. Here is a set of - and unfortunately I apologize, this needs a little explanation. Let’s say this is a 9th grade geometry class, and this is a quarter – one quarter – and this teacher has in the entire quarter given 5 summative assessments on which grade will be based. And this teacher has done a brilliant job of designing assessments that cover all of the learning outcomes. There are 6 learning outcomes. In this district you not only give a grade, you give student performance on ability to construct a defensible theorem or whatever. So, you’ve got 6 learning outcomes and they’re weighted for a grade, and you’ve got 5 assessments. And each assessment can be scored against all 6 outcomes, and they’re scored on a 1 to 4 rubric. So, one of them was a project and it was presented and it was scored on a 1 to 4 rubric. So, here are the 30 1 to 4 scores that

this student has accumulated at the end of the quarter. Now, in this school district you have to give this kid a letter grade for the quarter. And my question to you is "What's that grade?" And, again, I want you to talk to your neighbor. I don't want you to do any math – I'll tell you that if you add up all the scores it's 86 out of a possible 120, which would be a 71.7%. If you weighted them, using the weights, it would be a 76.5%. But I want you to figure out what grade the kid should have and then we're going to have a little discussion about what are the decision rules you used for giving that grade.

"The two that are 50% should weigh more obviously. And the two that have more weight are higher than the other ones...yes, I think you are seeing it exactly right."

"Right". "Here this child's lost in this learning outcome". "Yep". "OK, and here they fluctuate and here...". "So what grade would you give this one?" "According to this a C."

(MUSIC)

(MUSIC)

Okay, let's take first a show of hands. Who'd give this student an A? Who'd give this student a B? Who'd give this student a C? Okay, now imagine this grade book ends up in court and you have to defend the grade you gave? I mean, it can happen. What are some decision rules you used that would get you to either a B or an A? What are some things you'd say? Because again if you calculated this it's in the 70 percentile range – 71-77% - in most people's way of thinking that's a C. So what rules would you lay out that would justify a B or an A here? One of your grade calculation rules might be to weight later attempts. And you could come up with a formula if you wanted. The 1st one's 5%, the last one's 50%, sliding scale – whatever. What is another decision rule you could use to deal with the data in coming up with a B or an A for this kid? And it might help – let me back up – in this district you not only have to give a grade, you have to give an outcome score from 1 to 4 for each of the 6 outcomes. So you have to put on the report card a 1, 2, 3 or 4 for these 6 as well. What would you give the kid for 1? That's easy; it's an average, right? It doesn't matter how you do it – it's a 3. What would you give him for 2? Okay, another possible decision rule is drop lowest performance. You know, that's perfectly reasonable to say to students "hey, look, I know you're going to make a mistake – I'm going to drop your lowest performance in any assessment category before I total you." That would be easy. You drop the 1 and again the average works. What about #3? Who gives this kid a 4? Who'd give this kid a 3? The average here, of course, is 14 divided by 5 is a 2.8. This is what you want. By the way, this is the only learning outcome that as a teacher I'm proud of. This is the only one where the student obviously grew in my class. Now, what a lot of people do – one grading rule that people apply is to use the

most recent 3 pieces of evidence for the particular outcome. And if you did that, you'd have these 3, and the average of those 3 is going to be close to 4. Another thing you could do – you do this already – is either round up or round down. So you could round up. So if you applied – by the way, if you applied most recent scores, then weighting later attempts is superfluous, right? So if you applied these three – drop lowest performance, most recent scores, and round up – let's look at what you get. 3, 4, 4, 3, 4, 2, right? Then if you apply the weighting formula, these two are 4, so 50% of the grade is a 4 – I'm not going to bother to do all the rest of the math here, but simplistically the average of the others is in the 3 range. So, whether this ends up being a B or an A depends on a lot. But this is a perfectly valid way to grade. You could defend this in court if you had to. You don't need to average. And I'll tell you something else. You are going to have parents who are going to have issues. Parents are always going to – and students are going to have issues – they're always going to have issues. But there's a lot of research in psychology and in organizational dynamics and everything else that says the issues that people have are not with the decisions you've made, it's with your process. So if you communicate your rules up front, at the beginning, you've got no problem. You're always going to have someone who is going to disagree with anything you do, but in general you're going to have no problem. You can't go back now for this quarter and change your rules without creating issues, unless you haven't communicated anything about how you grade. One of the problems you have as teachers is that most of you were more successful in school than the average person. And people who are successful in school have a harder time questioning the assumptions that you lived and succeeded under. So my basic point is you need to have rules. You can sometimes get away with an early elementary grade levels not having rules. But you need to have rules – I would encourage you to have rules – but the rules can be whatever makes sense in terms of your grading philosophy. You obviously have to figure them out for yourself. But you do not have to average. By the way, when you do use numbers consider – medians actually are rarely a good idea, unless you're trying to think about this kid, because medians emphasize the central tendency – they don't emphasize growth. But the mode – the most common score is a decent approach – I mean the mode here is a 4. Let's move on to building a standards-based grading system – and I would argue there are 4 steps to building a standards-based grading system. The first is to build an assessment pacing guide – not a lesson planning pacing guide. Because if you take standards seriously the first thing you need to know is, how am I going to know whether the student has reached the standard? Therefore, what I care about is when do I teach what, but when should I expect the student to be able to demonstrate mastery of what? So build an assessment pacing guide from academic standards and course expectations. Second, outline an assessment plan. What

assessment types are you going to use? What's your balance between formative or diagnostics, and summative or evaluative assessment? What's your sampling plan? Because I hope you're not scoring everything. One of my favorite stories – it's probably hypocritical, but a president had a national security advisor, and asked the national security advisor on a report on a particularly important international issue. And the security advisor got the whole national Security Council working on this report and put together this report and put it on the president's desk and the advisor meets every morning with the president. The next morning the president held it up and said "Now is this your best work?" And the security advisor was horrified, and said "Sir, I think so. Let me take another shot at it." So they made everyone on the council cancel their vacations, called in people from defense, CIA, the embassies, they were really going to crank through this thing, and put it back on the president's desk the next day. "Now, is this really your best work?" Oh, my God. Just burned the midnight oil, exhausted every, every, every piece of intelligence available anywhere in the government, and got it back to the president. Same question next morning. Finally the security advisor – exhausted – said "Mr. President, if this isn't my best work, you have my resignation." The president said "Great! I'll read it." And you know what? That's what you want from your kids, right? You don't need to see everything they do. You're not the worker – they're the worker! So you need a sampling plan. You're not going to get that kind of response, but you need a sampling plan. You need reference tasks – tasks that you didn't create – that other people created, and you have some sense of how students should be expected to do on them, because there's evidence of how other students have done on them. Because otherwise all you've got is the reference set of your own classroom, and you know one class can be stronger than another one from year to year, and your own standards can migrate toward what you're finding. You need external reference tasks, and I'll come back to that. Timing, task-weighting – you know you need an assessment plan, and we'll take a look at what that means in a second. You need to share your plan – your record-keeping system and general principles with the students and the parents. And finally you need to analyze the results, you need to communicate, and you need to be okay with adjusting as you go. One of the most powerful things I ever did with my students is to stand up and say "You know what? I made a mistake. Here's my mistake. Here's what I think I should do about it. What do you think?" Now these are high school students, but you can do that with younger kids too. And the pay-off in terms of when I felt I had to stick to my guns was enormous. So, building assessment pacing guides. Now in Cleveland, actually we've done this and what you see here is topic-by-topic the students – the first topic of the school year is place value – numbers, numbers sense and operations. You see a copy of the Ohio state standards relevant here.

Use place value structure of the base 10 number system. You then see the next – the standards-based report card outcome. And you will see, uses place value to read, write, and represent and compare numbers. It's right here – it's the first thing in the pacing guide. Then next to it you'll see assessment N-2. In Cleveland, we've built 24 short-cycle assessments that cover the entire year of 4th grade math. And they're not benchmarks, they're not these sampling of items from across the curriculum that help predict a test score. They're actually deep, short-cycle assessment tasks that say "Okay, for number, number sense and operations, here are the definitions questions, here's an understanding question, here's an application question, so that you get back data that helps you understand how deeply do students understand this stuff. Then you see the textbook correlations and if you'll note the page numbers are all over the place, because the textbooks are not standards-aligned. We didn't even realize that, as a district, until we built these things. Then finally you'll see a suggested timeframe. The recommendation is your class probably is spending 4 or 5 days on place value. So, if you have this in Cleveland, it's fairly easy. You can go through this – this is in chronological order, and you'll understand that for the first quarter, these are the 4 report card outcomes that you're going to have to report on. This one, this one, this one, this one, this one, this one you're going to put not applicable for the first quarter, because there's no evidence – you're not going to create any evidence in your class for how students do on those four. And the report card gets printed out with explanations for that stuff so parents understand. So these are the four that you're going to have to report for the first quarter. Then the question becomes how am I going to assess that? And so what my second page does is it says here are those four outcomes, and what assessment tasks are we going to do together this quarter together with the kids. The first assessment in my class is day #1. It's a diagnostic assessment – number, number sense and operations. I want to know where these students are, using our diagnostic assessment. It's a formative assessment – the results have nothing to do with their grade – that would be grossly unfair. But I'm going to use that data to figure out what I need to know and what I don't need to know. Now I've got an assessment task – it's sort of an introductory quarter project called mathematical inventory of my world. It's a set of questions, they're supposed to do some drawings, they're supposed to do some calculating, they're supposed to write some stuff. And I've got a scoring report for it that I give them with it and I'll score it and give them feedback as many times as they want during the quarter, so it's both formative and summative. It really only deals with these three outcomes. It doesn't deal with this one. But, whatever they've given me – the best work on this they've given me at the end of the quarter, I am scoring and putting in the grade book and that's absolutely appropriate. Then, after we've taught place value and estimation, I do have an on-demand test – an assessment.

And that's appropriate. 4th graders have to take the state test – on-demand assessment is part of what they need to be able to do. It's totally appropriate to give a test and to score them – I'm okay with that. They also have a real world problems independent project that deals with outcomes #1 and #3. Another short-cycle assessment, an on-demand one, covers these three outcomes. A math journal sample – they have to keep a math journal – and this is my point about sampling. You don't have to score them all. There are a lot of different ways you can do it. You can collect them periodically, score a few entries, and just note that you've seen all the others, or you haven't seen all the others. You can have them show them to each other, score one on an overhead, so everyone has a sense of where you're going. There's lots of stuff you can do to get feedback. One approach might be, again that's all formative, at the end of the semester give me your three best entries and I'm scoring them. And that becomes part of the grade. Another short cycle assessment. A diagnostic assessment on variables and algebraic concepts for the second quarter. I'm going to give it in the first quarter, so I can prepare my lesson. It's not part of the grade. Again, that would be unfair. Then they do their student presentations on their independent projects. It definitely deals with outcome #1; depending on what they present I might go by the evidence on the others. And then finally, an end-of-quarter exam. In that end-of-quarter exam I would strongly encourage – as well as throughout here – use reference items. What do I mean by that? For example, NAEP reference items. Raise your hands if you know what NAEP means – a few of you. The National Assessment of Educational Progress – the nation's report card – it's the test that is used to say where Arkansas stands against the rest of the country. Now these items – and NAEP is not the only source of released items. You have released items from your state assessments on the web – you have released items from other states on the web – you have a variety of companies who produce this stuff. What I would argue you're looking for are assessment tasks that are related to your subject that are aligned to your standards, about which you have some data of how students are supposed to do on these, because that gives you – not proof – but a reference point to how your work relates. Let's take for example this one. This question – Laura wanted to enter the number 8375 into her calculator. By mistake she entered the number 8275. Without clearing the calculator, how could she correct her mistake? There's been a section on number sense properties and operations and this says what this question is related to – and this is the expert judgment of the people who developed this assessment task. You can agree or disagree with it, but if you disagree with it just don't use this. But it says what this question is supposed to be about and you can see how it relates to our topic. Then they provide information as to how kids did on this question. Now in NAEP, for those of you who don't know, proficient is a very high standard. There are probably only one or two states

in the country that define proficient in 4th grade math as rigorously as NAEP does. So if your kids are proficient against NAEP standards, you can feel good about that. And what it says is that of the students who are proficient – I’m sorry, there’s one other thing – there are 5 different levels of scoring on this question. The student can get an extended score, a satisfactory score, a partial score, a minimal score or an incorrect, off-task score, which is a 0. What you see here is that of the students who were proficient overall on the test – which is a high level of expectation for 4th graders – 24% of them got an extended score, and 31% got a satisfactory score. So if you were to put this question in your test, after having taught to this, and it comes back and 75% of your students you scored as either extended or satisfactory, that doesn’t tell you for sure that your students are above NAEP standards – it’s not statistically valid to say that. But it gives you a sense – you’ve asked the questions from outside your own set of assessments, and it gives you a sense that “Gee, I think my students are answering the kinds of questions that on a national scale would be answered this way.” Because otherwise, what do you know? You can use reference items, and a lot of schools have shared assessments – everyone in the school uses the same assessments – that’s wonderful! Even then, I’d want to make sure that there are some assessment questions in there that are from district-wide, or even better state-wide or national assessments. You know, Tims also does this stuff too. You need reference points outside of your own experience. Otherwise, your standards – your expectations – drift towards what you’ve got. And sometimes you’ve got way better than those expectations and you’re not being positive enough to your students – you’re beating them up, when in fact they’re wonderful. And you can beat them up, but you should also tell them they’re wonderful. So, that’s an assessment plan. You then need to keep good records of student results. And the problem with standards based assessment is there are three dimensions to this. If you’re a secondary teacher, you’ve got 150 kids. You’re used to your 2 dimensions – you’ve got your students, you’ve got your assignments. And when you’ve got an assignment that should probably be assessed against a couple of different standards, you’ve got a 3rd dimension. Now if you’re an elementary school teacher, more likely than not, you’ve got a single page for every kid, maybe even every kid every subject matter anyway. But for most people, the fact that you need to have assignment description information, the standards outcomes assessed, and student-specific results, means you have 3 dimensions. And that means that generally that’s more than regular grade books easily handle. Electronic grade books are strongly recommended, particularly for secondary folks. Unfortunately many electronic grade books aren’t very good at standards. I’ve been working with one district that has this student information system with this grade book. And the grade book was cooked up by some technology person who doesn’t understand standards, and

there's no place to assign multiple scores for particular assignments. Now you can get around it. You can take an assignment that you scored 4 different ways and enter it as Assignment 1-1, 1-2, 1-3, 1-4. But it's a little bit more of a pain. Here are some grade books I am aware of that people have said have good standards grading capabilities. I don't swear by any of them. I've not had the time to investigate them. I only offer them as some suggestions if you're really hunting for a solution you can Google and look around. Finally, you need to share the plan – the records and process – in an age-appropriate way. I've seen people take the kindergarteners and 1st graders and help them turn the standards into “I can” statements. And the students can keep a little checklist of things “I can do”. I've seen kindergarten teachers use the graphing as part of their curriculum anyway, to have the kids graph various aspects of their assessment results. Make examples of student work available wherever possible. I used to take an assignment that they were going to work again, but take an anonymous form version – sometimes I typed it up to hide the handwriting – and then score it on the overhead, and then tell them all “You saw what I just did to that one – now go work on yours again.” Build assessment rubrics and other scoring guidelines with students. Have them write a rubric. Have them score something against a rubric. Share feedback on scores with students. Encourage them to keep their own records. Explain the process and principles, and where necessary discuss. I don't think any time spent helping a worker better understand the goals that he or she should be working to is time wasted. If the student, as the worker, doesn't understand what work they're supposed to be doing, they're not going to be as effective or motivated. So, time spent discussing this, saves time in learning. So, in general, grades must relate to academic standards and course expectations. Public criteria and student work examples are reference points for grading. Grades should only be based on individual academic achievement. You can have group work, but you ought to assess individual achievement against the standards, because that's what the grade is. Sample student performance – do not feel the need to score everything. Keep records that can be updated easily. If you're not using an electronic solution, use a pencil. Crunch numbers carefully, if at all. Sometimes you just have a decision role, and you don't actually need to calculate percentages. And most importantly – and I'd love to teach a whole other class on using quality assessments – because most of us were not taught how to do assessment well, and that's really the most important thing. A bad assessment is not going to give you good information. But I want to deal with two more issues today. One is this issue of progress versus achievement. And then the other one is this whole question of keeping anchored in a sea of testing. Progress versus achievement – I don't have any good answers here. I'm going to tell you what Cleveland did. I'm not sure I'm going to tell you it was a good idea, because we created a fairly complex situation, but it's a fairly

complex problem. We sat down early on to create a standards based grade book. And, of course, if you believe in standards you don't want to give grades at all – right? Well, that's a non-starter in our community. We have a large percentage of parents who I think of the beer mug parents. Tell me how full my student is – fill him up and tell me how full he is. And we knew that if we got rid of grades, we'd never get a standards based report card out there. So, what we did was we kept the grade. We also have these standards based outcomes, and you saw the 4th grade math one. And we said we're going to put the standards based outcomes on the inside. People entered the grades electronically and we printed out an 11 x 17 piece of paper that folds into a 4 page booklet. And the front page has the grades and the behavior marks. You open it up and there's the report card outcome. And our goal – we figure right now we've got about half of our parents opening it up. I'm not happy about that, but it's the price that we pay for having a standards based report card. But our goal is to increase that to 100%, and ultimately get every parent comfortable enough that we can get rid of the grades. So, we kind of took an evolutionary approach to that. So, we then had a problem. Our standards based outcomes are end-of-year achievement indicators – 4 means the student is exceeding the standard as defined by end-of-year expectations, 3 is meeting the standards as defined by end-of-year expectations, 2 is approaching, 1 is attempting. So, it's perfectly conceivable that 1st quarter a student should have 1's, 2nd quarter 2's, 3rd quarter 3's, 4th quarter 4's. And the right standards answer to that is so they could get a D 1st quarter, a C 2nd quarter, a B 3rd quarter, an A 4th quarter – and that's all that matters because it's an end-of-year grade anyway that matters. Right? Wrong! And one of the more interesting injections that came up for me is the athletic people said “Well, we will have no fall sports team,” because we have to calculate the GPA on a quarterly basis. So we said alright – we're not willing to build quarter-by-quarter expectations because we don't have enough clarity in our curriculum right now to do that. So we're going to have end-of-the-year expectations on the course outcome, and we're going to build these grades as is the student making adequate progress at this point of the year, so that student with all 1's in the 1st quarter might get an A, all 2's the 2nd quarter might get an A, all 3's in the 3rd quarter might get an A, but in the 4th quarter, in order to get an A, it better be all 4's. Incredibly complicated. In our first year in going district-wide, it actually worked on its face. These bars are the distribution of grades district-wide – it's about ½ million grades every quarter – a big district. So here's the percentage of F's 1st quarter – by the way F for 1st graders is not an F – it's an unsatisfactory, but it's the same thing. Here's the percent 2nd quarter, 3rd quarter, 4th quarter. And you can see the distribution was fairly consistent every quarter. So on the face it looks like our teachers got it. They understood that students might consistently get B's all year long, because it was achievement appropriate

for that time of the year. And then for the course outcomes – and there are about 6 outcomes per grade – here was the distribution of outcomes in the 1st quarter. Again, it looks on the face of it like teachers got it. Most of them, the mode was 1. And then 2nd quarter you see growth. This is about 3 million outcomes for each quarter. 3rd quarter you see more 3's. 4th quarter you see more 4's and the mode is now 3, which is meets the standards. So on the face of it it looks pretty good. It's not that good. Here's an actual report card printout for an individual student in the district. I just picked a couple randomly to show you. 1st quarter the teacher gave the kid all 1's. Now I already showed you I think only these 4 should have been scored in the 1st quarter, so I'm curious whether the teacher really covered all these things. Secondly, I'm amazed that they're all 1's, so I'm a little confused about how that happened. If that was all bona fide – what's nice about it is all these ones could be an A, then 2nd quarter all 3's and 4's could be an A, 4th quarter all 4's could be an A. So it looks like this person got that progress versus achievement issue, but I'm skeptical about where these scores came from. Here's another kid's report card. The teacher appears to have had some non applicables – that's good, that's the right answer– all 1's was a C. I'm not sure why all 1's here is a C, but all 1's here is an A. But I know that we've got work to do in terms of helping teachers all get on the same page about what are appropriate expectations. Secondly, I'm okay with the 1's here resulting in a D, because if there's been no growth – again our model says the grades should slip. What I'm having trouble with is that. Now, in the terms of our culture of our district I have a pretty good idea what happened to that one, even though I never talked to that specific teacher. The disconnect here is stark. In the terms of the culture of our district, there are a lot of people who give kids F's in the 3rd quarter as a shot across their bow. And a couple of people in some other sessions have said “How can you do this? You could create all this mess.” You know what? The mess was already there, and that culture was already there. All we've done is expose it to the light of day. Now, I blame myself for this because there's a lack of professional development in helping teachers come to grips with this. Our decision was to get it out there, get the data flowing, use the data to push the professional development agenda. But this teacher clearly doesn't get it. And what's probably going on – homework, attendance, behavior – I'm going to tell this kid he could fail – not a standards based mindset. So this is what we've done in Cleveland. We've got a lot of work to do. But, the interesting trade-off between year-end outcomes and progress appropriate to the time of year is interesting. And I'll tell you something else that we dealt with here. Special education – before we put report cards on-line – the 1st year we put grades on-line we got a flurry of calls from special ed teachers. I've got a 4th grader and I want to put A/2 on his report card for reading. He got an A at the 2nd grade level. Well, anyone who knows the special ed laws knows that that's

a violation of law. You can't have a report card that so clearly demonstrates that this is a special education student. So, we said you're not allowed to do that. And the special ed department was horrified. They had no idea that practice was as pervasive as it was. Because it was all hand-written, they never saw a kid's report card. So we drove out one bad behavior, but with this issue, the standards answer for that kid who's in 4th grade doing math at a 2nd grade level, who gets all 1's on these things, and he's a special ed kid with an IEP for math, is that he gets all 1's, because he's being assessed against the 4th grade standards. You may not like that answer, but that's pretty clearly what No Child Left Behind says, and that's pretty clearly what standards based education says. Students should be assessed by the grade level in which they are enrolled – the expectations in which they are enrolled. And again, if you take standards to their natural conclusion, it doesn't bother you. Because the issue is time isn't fixed, results varied, results should be fixed and time varied. So this kid may want to spend a couple of years in 2nd grade. Let's get rid of the grade denominators, let's just focus on the outcomes a student needs to achieve. But obviously that's not the way it works. So what we worked out is if the kid's in 4th grade and you've got an IEP saying – 2 things – one is you can assess alternately – just fine – that might allow you to put a 4 where you might not otherwise be able to demonstrate a 4. But if in fact even the alternate assessments show 1, but the IEP says that's good for this kid for this point in his development, the IEP might give you permission to make these all 1's and give the kid an A. And then the legal people need to look at that to make sure that that wasn't again making it clear that it was a special education report card, and the legal people basically said “That's fine, because this is so confusing no one is going to figure that out.”

(MUSIC)

(MUSIC)

Standards based education is very clear. You report achievement against the standards. If you want to show growth, you need to find a way to do that, but it's not a standards based report. You might have an achievement and a progress grade. You might take the standards and break them down so you say here's what's expected at the end of this grading period, here's what's expected at the end of this grading period, so you don't have our little mess of the 1's, 2's, 3's and 4's. We weren't ready to do that yet, so we took a short-cut and it's costing us. But you do need to deal with this issue. Finally, I would ask you to – as you deal with more and more testing – to please protect yourself from the experience of the Chicago classroom teacher. A survey of classroom teachers in Chicago, when they introduced the high-stakes Iowa test of basic skills for promotion a bunch of years ago, over 2/3's of Chicago teachers reported narrowing their curriculum and

lowering the high order thinking skills that they were addressing in the classroom – exactly the wrong thing to do if you want to improve test scores. But if you had no professional development and support and you had no accountability guidance to the contrary, drilling and skilling is a natural reaction to high stakes. It's incorrect, but it's natural. I would argue that an important way – that what you ought to be focused on are standards. Communicating clear expectations, working to help your students get to those expectations. But you're going to need to armor yourself in the coming standards based world – in the standardized testing avalanche that's upon us. One thing we've done in Cleveland to try to help teachers arm themselves is to give them feedback on how their grades correlate to test scores. And we don't expect perfect correlation. I would hope your grades are much richer and more nuanced assessment of students and how they do on test day. But, just to give you an example, we used to give the 4th grade state reading test in the fall, right around the end of the first marking period. So, students were given an advanced or proficient or basic or below basic rating on this. It just so happens, that the Ohio state reading test is proficient is roughly basic for the NAEP test, and our standards are aligned to the NAEP standards, so it turns out that approaching the standards, a 2 is roughly comparable to proficient on the state test. So it was our belief if that you look at our 4th grade reading outcome, the course outcome comprehends beyond literal understanding to examine, predict and occur is the dominant outcome that is actually being assessed by that state test. So we said "Look if a kid is getting a 2, 3, or 4 on that outcome on his report card in the 1st quarter, he should be proficient or advanced on that test." And if the kid is getting a 1 or a 0 – not applicable – then the student should be basic or below basic. So these gray areas are the areas we say should be filled in. It doesn't mean there's never a kid up here or a kid down here. I mean, a kid could have a bad test day. But the bulk of your kids should be here, and district-wide we found 70% of our kids were there. What's interesting is, some teachers – 20% of their kids were there. They had no correlation between test scores and report card outcomes. They need help. Also, we find many more kids here than here, and that's common in lower performing school districts. In lower performing school districts and lower performing schools people tend to give students grades and outcomes that are higher than their real achievement. It's understandable, you feel bad, your standards slip, you don't have those reference scores to understand where these kids really are. But the problem is you're sending mixed signals to these children. You're not telling them what they need to know to improve to get their scores to standard. So, in places where this goes on, test scores tend to be lower. That's one thing we've done, another thing we've done is to say look at your course outcome. On the 3rd grade reading test which is given after the 3rd grading period, kids had an average of 3 or more on their report card outcomes, 70% of them were proficient

on that test. Because 2 or 3 41% were proficient on the test and 2 10% were proficient on the test. So I would encourage you if your district doesn't give you this kind of data, or your school doesn't do this – you know once year sit down and come up with some numbers like this. They will help you have confidence in your grading system so that you can put away the pressure of the high stakes test and focus on what you know matters. And when you get those parents and administrators – and students - in your face about test prep, you can trot out stuff like this and say “I hear you, but trust me. We've been through this before. I know something about the relationship about the feedback I give in class and how students perform on these tests.” And I would argue that even if your results are not statistically significant they can give confidence to people – including yourself – who are struggling to stay focused. Because otherwise you're going to end up, many of you like Chicago teachers. So what I hoped to do today was kind of begin to explore the implication of standards, discuss some practical ways to integrate standards based instruction. Look at this question of how to stay focused high stakes testing, but fundamentally I know that I have not taught anything today. I hope what I've done is provoked you to think about where standards ought to be taking your grading and reporting practices in your effort to make the students the workers. If I've done that, then I hope I've connected you to some materials that you can use to begin to have conversations with your colleagues. Just doing the parachute-packing thing – that's something you can do in a 35-minute grade level meeting and really get a great conversation going. Finally there are some resources that I'm aware of that I like, and there are others I'm sure that many of you know. Someone came up to me and said Thomas Gusky has written very well about standards based grading. Lauren Rusnick at the Institute for Learning – she's really one of the founders of the standards movement, and they're very articulate about high expectations for students and the fact that smart is not something you are, smart is something you get by working hard, and we have a culture that needs to change our whole attitude and it's very helpful. Stiggins and the whole issue of what high quality assessment is is of course central to this, because if you don't have high quality assessment, what are you scoring and reporting? And then finally Ken O'Connor is where I get much of this stuff about grading, but Marzano has also done a book on this and Gusky also done a book on this. Thank you and....

(MUSIC)

For more information or a free on-line follow up to this program, log on to www.ed.gov/teacherinitiative. This broadcast and the follow up are brought to you through a partnership of the US Department of Education and the Panhandle Area Educational Consortium.