# TRANSCRIPT OF PROCEEDINGS 

BEFORE THE
COLORADO DEPARTMENT OF EDUCATION COMMISSION DENVER, COLORADO

May 13, 2015, Part 5

BE IT REMEMBERED THAT on May 13, 2015, the above-entitled meeting was conducted at the Colorado Department of Education, before the following Board Members:

```
Marcia Neal(R), Chairman
Angelika Schroeder (D), Vice Chairman
Steven Durham (R)
Valentina (Val) Flores (D)
Jane Goff (D)
Pam Mazanec (R)
Debora Scheffel (R)
```

MADAM CHAIR: We had to postpone the discussion this morning regarding $C$ Mass, and so we are now reconstituted with the C Mass discussion. Commissioner.

UNIDENTIFIED VOICE: Thank you very much. This item is brought to you again. We try to bring back to you is two options besides the original one that we presented as we feel they're required by law to set and cut scores, and that's the responsibility of the Board. We've tried to give you some options here for discussion today. So with that, I'll turn it over to Joyce Aposki (ph).

MADAM CHAIR: Thank you. Joyce.
MS. APOSKI: I've had some opportunity to think since earlier today, and before we jump into what the options are, $I$ want to take this opportunity to talk briefly. I promise briefly about norm reference testing versus Criterion referenced testing.

So when we had a Criterion reference test, we are assessing against a set of expectations, a set of content standards, and a set of performance levels. What we're looking at is what kids know, and are able to do against those pre-established standards, those preestablished expectations. When I have a Criterion reference test, $I$ can score that test based on a single kid, because I am scoring that student against those content standards, not against any other kid.

For a norm reference test, we are assessing students against how other students perform. So the score I get is dependent upon the other students who actually test that day. It is not connected directly back to a content based decision. It's a distribution of scores. So in the end, with just a percentile rank, I cannot answer the question of what it is that the student can, and cannot do.

As a matter of fact, $I$ could have a test where just about all of my students really actually do meet the expectations, and yet there will be some students who score at the second percentile. Or I can have a test where none of the students meet the expectations, what we actually want our third graders to be able to do, and yet there'll be students who are scoring at the 98th percentile. It's a ranking system.

Kids who know the least, to kids who know the most, but there's not a connection back to the standards. Both our state law, and our federal law, and those systems are based on the premise that we will have a Criterion based system.

Okay. Something that is generated off of the standards, which is why we started first with the content standards, then developed those performance levels, and now we're trying to establish cut scores based on those
performance levels that's based on the content. It's a content base decision.

If we were going to establish scores that were based on more of a norm reference system, what would happen is that rather than pre-establishing what it is that we expect kids to know and be able to do, and measuring the kids against those expectations, we would actually be setting the standards. What do we want kids to be able to know and do based on actual performance? It would be the exact opposite of what the system is set up to do.

When we're looking at standard setting expectations and score requirements, again, under both state and ESEA, there's reference to being consistent with relevant nationally recognized professional and technical standards, having adequate technical quality, and meeting the requirements under the law.

We have to report out on performance levels. We have to have those performance levels be reflective of proficiency. The results have to include scores that are aligned to the standards. They must include performance levels. They cannot just be raw scores, scale scores, percentiles, things like that. The school and the district reports that are generated have to reflect those performance levels as well. One of which has to reflect what we mean by proficient.

So when we're looking at normative data or descriptive data, what role does that play? We, in the generation of our individual student reports, took very much to heart the concept of we are moving to brand new standards that, frankly, schools and districts are in the process of implementing, and it's new to students. And we knew from the Criterion based perspective what the standards were asking our students to know and be able to do, it was going to be tough.

So we wanted to provide additional information that would allow folks to also ask questions of, okay, so my student may not know everything that they're supposed to know according to the standards, but how do they compare to other kids, to other schools, to other districts?

You had in your packet two sample reports, and we had been requested to bring those reports with us today. And when we look at those reports, I want to point out where we're talking about, kind of a Criterion based system, versus where we're talking about where you can make some of those other comparisons to other students.

So at the top of the report, what we have is, obviously this is a content area for science. There's a scale score. The scale score is what allows us to compare performance from one year to the next year. So
here we have a scale score of 573 , that has been assigned to a performance level. And it's of moderate command. The students diamond is at the top. It's the big diamond. And when we look at that student's diamond, yes, it is sitting at that moderate command level. When we compare that student's performance to the student's school and district, that student is actually a high performer. Right. They're doing better than the average for that school or district. But when we compare that to the state, the student is not doing as well as the state as a whole is.

And again, that's one of the purposes of the state assessment is to be able to look outside of just your community, and have a broader perspective. So again, this student has a moderate command of the standards. The student is doing better than the school or district on average, but the student is not doing as well as the state on average.

We then break that score down into sub scales, and for science we have physical science, life science, earth system science, and then scientific investigation, and the nature of science. For social studies, we have history, geography, economics, and civics. With this, we can start to look at a deeper level in terms of how the student is performing, and when we look at
physical science, again, the student is performing better than the average for the school or the district.

When we look at life science, the student is performing slightly better. When we look at earth systems slightly worse, and when we look at scientific investigation that does appear to be a strength area for this particular student.

As a teacher, I'm going to be especially concerned about the earth systems piece. Right. Why is my student not performing as well on that area as the other areas? Why is my student not performing as well as his or her peers within the school, within the district?

When I look at the next part of the report where we see the earth systems breakout, we have the grade level expectations, and we can see that the student again is at the top. Then we have district average, then we have state average. And where we see the biggest difference is actually in GLE 2, which deals with the earth's surface changes constantly through a variety of processes and forces.

So we've narrowed down what it is that this particular student seems to be struggling with. The student in terms of weather, is solid. Right. The student is actually doing better than his or her peers within the district, and is even doing better than the state average.

Again, all of this is comparison of student to student, something which I've heard referenced to from this group before. Wanting more of that, less of the reference back to the Criterion.

We also have the scores broken up into what we refer to as selected response. In the olden days this would have been multiple choice. They look slightly different in a technology enhanced world, but for our purposes we can consider those multiple choice.

And then constructed response. And those are the items where students actually have to create their own response. They're not just selecting. They have to actually write it out. And when we look at the student, the student is performing pretty consistently with the school in the district on a selected response a little bit below where the state is.

When we look at constructed response, the student is actually outperforming the average for that school, for that district, but is still a little bit below the state.

This report is actually an example from fifth grade, because fifth grade already has the approves cut scores, so we were able to complete the reports by having this very first part where we have it split into the four performance levels. And what we are asking for you to
do, moving forward at some point, is to give us cut scores so we know where to place those four levels. The rest of the reports are done.

We also know that when we were going through the standard setting process, we did also look at student performance. It was not done in isolation. Our standard setters very much came from the perspective of content. Right. What do the standards demand. But we did share with them what performance looked like. How many students got the items correct? We did share with them distribution, and we asked them to explain the distribution. Anything that was different than what they would have expected based on their own experience.

The standard setters, as they indicated back into March, felt that actually where those cut scores fell and what that distribution looks like, was matching their experience within the schools and the districts. We've been asked whether or not there are other standards setting options. Again, we're coming from the perspective of a Criterion based system.

I did provide for you in your packet, I think a description of a couple of the other models. So, yes, we used bookmarking. We could have used (inaudible), we could have used a modified (inaudible), but in the end they're all content based, and they all in part our
dependent upon expert opinion, educators opinion, and there's no reason to expect that had we followed one of those, we would have necessarily ended up with results that looked very -- sorry, very different from what we actually have. Again, it's the standards that are driving where these cut scores are.

After the March meeting we did consult with some of our national experts about what our options might be in order to meet your wants as well as still be technically sound in the end, because we know that we want to have a technically sound system, both for our state purposes, for our federal purposes, and just good practice.

The (inaudible) that we had followed, a technically sound and nationally recognized process. Okay. This is not something that we just made up. It is technically sound, nationally recognized, but they did agree that there was some room to adjust to the cup clients, but we would still be able to defend from a technical perspective. And that was our goal, to be able to make an adjustment, but also be able to defend from that technical perspective.

You have those adjusted cut scores in your packet. They are not radically, radically different. We look at the March recommendation, and we had about 19 percent of our students who were scoring at strong and
distinguished. If we would make an adjustment just to the strong cut, which was one of the options that we have, we would have about 25 percent of our students in strong and distinguished. That stays the same for option c, but what happens with option $C$ is we also adjust the moderate cuts, and so we have fewer students in unlimited category.

So with our option C, we have 29 percent of our students in the limited category. Under the March recommendation, there was 29 percent. This is for science. For Social Studies we had approximately 10 percent of our students in March scoring at the strong, and distinguished mark. With an adjustment just to the strong cut, we have 16 percent, a split of 15 and 1 . And then if we adjust all levels it actually shifts to 14 and 2 . And again, we see a decrease in the students in the limited area dropping from 44 to 36 percent.

Do we have some external markers to gauge the reasonableness of these cuts? Back at the March meeting Mr. Jeremy (ph) asked, "What about AP? What about those advanced placement courses? And at that time I did not have those numbers off the top of my head. We do know that in Colorado we have about 14 percent of our graduates who earn a three or higher on a history or social science assessment. For college credit students need to earn a four or more. Okay. So that's three does not get a
student college credit, but again, that's about 14 percent pretty consistent with what we have here at 16 percent.

When we looked at the science advanced placement, we had about 7 percent of -- 7 percent of our graduates, um, earning a three or higher on a science assessment. Keep in mind for both social studies and science under advanced placement, there are a variety of tests. They're not comprehensive, so students can take just biology or just chemistry.

Also, we know that within the last month, NAPE released their eighth grade U.S. history results, and 18 percent of students scored at proficient or above. Again, it's a different grade level, but it is an external marker to just kind of give us a gauge in terms of is this kind of falling within the realm of what we would expect. And it's not radically different. And we know that these scores, especially in the social studies come more in line with what we saw with elementary and middle school.

So, questions?
UNIDENTIFIED VOICE: Yeah.
UNIDENTIFIED VOICE: Thank you. So can you just tell me if this sequence is right? And this is why the cut scores are problematic in my way of thinking. Is it right that we, of course, develop language that represents the standards in Colorado for social studies,
and science? That language exists on our websites. Then performance level indicators were identified as subsets or distillations of all that language, because when we pull up that language it's voluminous. We could read any one of those bullets, and kind of walk away asking ourselves, what did it say again? Right.

I mean, because the -- when one thinks of standards, one often thinks of very targeted, succinct, detailed language with verbs that are descriptive. And when you pull up our standards as -- as -- as many standards, one has a hard time landing on the verb that says what exactly does this mean? But, nonetheless, we have that language.

Then the performance level indicators are sort of a distillation of that language that creates the -the weights of how many items are aligned with that content. Then the next step is to develop items that link to the performance level indicators and the standards. Then the next step is to get kids to respond to those items with varying levels of proficiencies. Then the next step is to create rubrics, that groups of people would sit down and superimpose on levels of responses on those items. And then the next step is to get focus groups or groups or whatever we're calling them, to sit down and, say, yes, this kind of a response looks like it represents
distinguished. This kind of a response looks like it represents strong. Now we will align the scoring of those items across thousands of kids with that sequence.

And -- and so if I'm correct in that sequence or roughly so, then -- then I -- I guess my feeling is when you're ending up with one and two percent of students performing at a distinguished level, it strikes me as an artifact of all the people that did the work. So can you respond to that?

MS. APOSKI: Madam Chair?
MADAM CHAIR: Yes.
MS. APOSKI: Can I make a couple of clarifications about the process that you outlined first? So you're absolutely right that we start with the standards, and the science and social studies standards were adopted in December of 2009, and keep in mind that those are Colorado. Right. We're -- we're not dealing with any external issues when it comes to these two tests.

You're right that we then also develop what I'm going to refer to as performance level descriptors. And what that does is that takes those standards and it breaks it up into different levels. So we have our limited command, our moderate command, our strong command, and our distinguished command, right. So we're going to break up whether we have a student who is performing at -- kind of
like the entry level for that standard area is approaching what we would call kind of proficient, whether we've crossed the bar, we're strong or whether we're at that distinguished level.

UNIDENTIFIED VOICE: Let me ask -MS. APOSKI: When it comes to --

UNIDENTIFIED VOICE: (Inaudible).
MS. APOSKI: Sorry.
UNIDENTIFIED VOICE: When I -- when I -- I reread the performance level descriptors that are kind of a distillation of the voluminous language in the standards, I -- I can't recall if the Board voted on that language. I know that when I looked at it, I thought, and I -- I remember talking with Jill at the time, and I said, "Where did this language come from again? Who wrote it? Who created these bulleted items that are supposed to be a distillation of that larger body of work? And did we vote on that or did we just review it?" I don't remember voting on it.

MS. APOSKI: Madam Chair.
MADAM CHAIR: Yes.
MS. APOSKI: I'll answer that question and then I want to jump back to one other thing.

MADAM CHAIR: Okay.
MS. APOSKI: Okay. The performance level
descriptors are actually one of the things that you're voting -- well, will -- you're being asked to vote on. And they're joined with those cut scores.

UNIDENTIFIED VOICE: Okay. Thanks.
MS. APOSKI: Okay. In terms of the items, the items are actually I would suggest written back to our evidence outcomes that are coming out of our standards, rather than coming straight off of the performance level descriptors. So just a slight adjustment there in terms of how far have we journeyed.

UNIDENTIFIED VOICE: Except that -- excuse me.

MS. APOSKI: Entered so --
UNIDENTIFIED VOICE: So performance level descriptors, though, are written based on distinguished, strong, moderate, limited --

MS. APOSKI: Yes.
UNIDENTIFIED VOICE: -- so they have a huge
impact on those cut scores?
MS. APOSKI: Madam Chair?
MADAM CHAIR: Yes.
MS. APOSKI: Absolutely. And that's what those performance level descriptors are supposed to do. They are supposed to help us orient, and gain an understanding of what it is that our students know in terms
of those skills and concepts that are represented in the standards. And be able to separate --

UNIDENTIFIED VOICE: And you haven't (inaudible) yet.

MS. APOSKI: Madam Chair?
UNIDENTIFIED VOICE: Thank you.
MS. APOSKI: That is what you're being asked to take into consideration today, and what you were asked to take into consideration as well as in March.

Another piece that I would just like to clarify is that when it comes to scoring our constructed responses, they are not scored from the perspective of limited, moderate, strong, distinguished. That only applies to a total test level. And so we have some items that are worth two points. So there's a zero, one, and two, but there's not a direct correlation to these performance levels.

And I don't -- I just had to clarify, because people get confused about that sometimes thinking that the performance levels are applied at that item level for that scoring, and it's not.

UNIDENTIFIED VOICE: Yeah. I'm going to a very rudimentary level, because I went through a master's program that was competency based, and you either were performing, you were competent or you were not competent.

And I went through two years where professors, it was just like this. They were creating the program. So I went through courses where I didn't even get a score, because they hadn't decided what, you know, what that rubric or what would look like. It was very -- oh, anxiety producing.

So as an example, in one course, which was science education, the end what $I$ had to do was create a machine that worked, and worked for three minutes. And I mean it -- it -- it was a machine that moved and flew. I just don't see how we have gotten from that to -- to this where you either are or you're not. I can understand that. But these -- these levels that are, I think, I don't even remember that in theory. What theory as far as -- I just didn't know that theory, and I really have kept up in this area.

I even went to Mexico, and helped the department. The you -- at the Mexican Department of Education with our states to develop programs in competency based education. And my description is what I just described. And it didn't have this $a, b, c$. We were trying to get away from $a, b, c, d$, and $f$, which is basically what you have in -- I mean, if -- if you gave these descriptors, you could actually say that level up -up above is an a, and then a b, and then a $c$, and then a $d$,
and an f. You have five levels.
MS. APOSKI: Madam Chair?
MADAM CHAIR: Yes.
MS. APOSKI: So as we're looking at this concept of competency, and I -- and if you would take what you had, which was essentially a two kind of level approach. Right, it was, you're either are or you're not. Right. What has happened with most data assessments is we have said, not only do we want to be able to distinguish the kids who have it from the kids who don't have it, but we want to be able to distinguish those kids who have really excelled from those kids who haven't. And we want to be able to distinguish those kids who really are extremely struggling from those kids who are approaching competency.

So if we were going to look at as an example, you just have, like, kind of like one item in life with your example, but you could have built it --

UNIDENTIFIED VOICE: (Inaudible).
MS. APOSKI: -- you could have built a
machine that flew, but it didn't run. I think you said it had to kind of run and fly. I don't know. But you could have done one but not both.

UNIDENTIFIED VOICE: (Inaudible).
MS. APOSKI: And you could have had somebody
who, frankly, at the start did not have a clue as to how to even get gears working versus somebody else who could get the gears working. And you want to be able to distinguish those. So what we can do on the state assessment is reliably kind of put kids into four different categories with three being, I've crossed that line into competence, but $I$ can also then distinguish my kids who really, really are excelling.

And then I have my kids who are below that line of competence, but be able to distinguish my kids who are just at the emerging level from the people who are approaching.

MADAM CHAIR: (Inaudible) conversation.
UNIDENTIFIED VOICE: I have another question that (inaudible).

MADAM CHAIR: Sure, what it?
UNIDENTIFIED VOICE: Yes, it's related. So what happens to those kids? Do they retake that -- the teacher re-teaches in a different manner, the material that was not? So what happens? They just don't pass? And how -- how is that going to work?

MS. APOSKI: Madam Chair?
MADAM CHAIR: Yes.
MS. APOSKI: In Colorado, we do not have our tests as high stakes for students. Meaning we do not have
requirements that our students have to pass the state test in order to, like, graduate. In some states they do. Right. You cannot graduate unless you pass the state test. We don't have that in Colorado. We're not high stakes for students.

UNIDENTIFIED VOICE: We are high stakes. High stakes for teachers.

MADAM CHAIR: Wait. Wait.
UNIDENTIFIED VOICE: I'm sorry.
MADAM CHAIR: You were dominating the conversation. No one has the right to do that.

UNIDENTIFIED VOICE: I'm not dominating.
MS. APOSKI: Madam Chair?
MADAM CHAIR: Yes.
MS. APOSKI: And so, and that's why I said we are not high stakes for students. Right. The students, their graduation is not dependent upon whether or not they pass the test. Will the state assessments be used down the road for purposes of educator effectiveness? The intent is for the state assessments to be used for purposes of educator effectiveness down the road. Is that happening in this year? It's not happening this year.

MADAM CHAIR: Yeah, I would just add to that. You've got a room full of students, all of whom have parents who want to know how their children are doing.

How, you know, you need to be able to give them some explanations, which is what you're looking at. And I -- I -- it seems like we go through different stages in education where, gee, we don't want any grades, because they might feel bad or they're not what, you know. I appreciate what you're doing anyway. So, go ahead, please, unless anyone else has questions.

UNIDENTIFIED VOICE: Well, there's questions over here.

UNIDENTIFIED VOICE: Let Angelika and Goff go.

MADAM CHAIR: Yeah, let's get Angelika and Jane in here.

MS. APOSKI: So if you want an example of high stakes for students, kind of high, the AP test, which also has what, five different levels.

UNIDENTIFIED VOICE: And we're talking about the park.

UNIDENTIFIED VOICE: No, wait a minute. UNIDENTIFIED VOICE: We're talking about the (inaudible)

UNIDENTIFIED VOICE: We're not talking by the park. We're talking about the C Mass.

MADAM CHAIR: We're not talking about the park at all.

UNIDENTIFIED VOICE: But what we were talking about is how Criterion reference exams are scored. So you've got five on the AP. I don't know about the AB -IB. The NAPE seems to me to be identical, and, in fact, it goes right along. I mean, the scores are right in alignment --

UNIDENTIFIED VOICE: Right.
UNIDENTIFIED VOICE: -- with the scores we're having here.

UNIDENTIFIED VOICE: Right.
UNIDENTIFIED VOICE: So for us --
UNIDENTIFIED VOICE: But (inaudible) concerned --

UNIDENTIFIED VOICE: But you're talking about is a pass --

UNIDENTIFIED VOICE: (Inaudible) the question $I$ asked before. The question I asked before was what about -- we want those kids to -- to get that knowledge, to have that knowledge. It's not about, it doesn't matter because, you know, so we -- we get a template or we -- we have an assessment that says these kids do not know that knowledge. So what are we going to do? Because it's important to me, and I think it's important to parents that kids have that knowledge.

UNIDENTIFIED VOICE: Agreed. But you're
talking about what we do with the scores and we are just trying to get to scores. So I don't --

UNIDENTIFIED VOICE: Now, see you're talking about a test. I'm talking about content and knowledge. I'm sorry.

UNIDENTIFIED VOICE: What?
UNIDENTIFIED VOICE: You are still keeping
to that test. And I'm talking about there's a body of knowledge to know. There's a --

UNIDENTIFIED VOICE: Right.
UNIDENTIFIED VOICE: -- there's content to know.

UNIDENTIFIED VOICE: Right.
UNIDENTIFIED VOICE: And I want those kids, even if they fail that test to retake that course, to have teachers re-teach.

UNIDENTIFIED VOICE: No -- no disagreement with you.

UNIDENTIFIED VOICE: Well -UNIDENTIFIED VOICE: What we're talking about though --

UNIDENTIFIED VOICE: -- but that's the importance of -- of -- of education and learning. That's what it's about.

UNIDENTIFIED VOICE: And that's the next
step after you -- after a kid finds out where he or she -UNIDENTIFIED VOICE: Yeah.

UNIDENTIFIED VOICE: -- lands on assessment.
UNIDENTIFIED VOICE: (Inaudible).
UNIDENTIFIED VOICE: Then we've got a job to do, but we're not -- we can't even get there. And by not giving any kind of feedback to the kids.

UNIDENTIFIED VOICE: Yeah, we could give them a score.

MADAM CHAIR: You're not listening.
UNIDENTIFIED VOICE: Well, that's what we're talking about.

UNIDENTIFIED VOICE: Madam Chair, could I (inaudible)?

MADAM CHAIR: Jane.
MS. GOFF: The tests that were given last fall, and what grades were those again? Seventh?

UNIDENTIFIED VOICE: Madam Chair? Sorry.
MADAM CHAIR: Yes.
UNIDENTIFIED VOICE: The tests that were given in the fall were actually given to the 12th graders. So they were pretty much a culminating assessment.

UNIDENTIFIED VOICE: What's been released?
UNIDENTIFIED VOICE: Madam Chair?
MADAM CHAIR: Yes.

UNIDENTIFIED VOICE: We have released the elementary and middle school science and social studies. We did that last fall.

UNIDENTIFIED VOICE: Okay. (Inaudible). That's what I -- I thought there was just a lower grade levels that people had seen the results. I've seen -- now, we've -- we've all seen an exemplar or a sample of what one of those reports might look like that would go home to a parent, but I can't remember right now is -- is there a score on there where we're literally -- I can't remember. It's in some number of points out of a certain maximum number of points. And then in addition to that there is --

MADAM CHAIR: (Inaudible).
UNIDENTIFIED VOICE: -- the bar --
UNIDENTIFIED VOICE: This --
UNIDENTIFIED VOICE: -- and a diagram type structure that shows where it play -- where along the scale of proficiency. If we're still calling that. Actually, it's competency.

UNIDENTIFIED VOICE: Madam, sorry.
MADAM CHAIR: Yes.
UNIDENTIFIED VOICE: Are -- are parents seeing that, and then the second part of that, I know they are. But so the second part is, has there been any feedback, direct comment -- comment, feedback questions,
the good, the bad, and the ugly about what parents, and -and families, and teachers for that matter, are learning from that information? What kind of conclusions are they drawing? And to me that's where the teaching, and learning, and instruction part of this kicks in.

It's also purely achievement level, and whether you want to call that a raw score or a proficient -- a competency -- used to that -- level, that's to me it's all the same thing. However, people are talking about this, and what they're all learning, how they express their own learning conclusions, observations, to me that's what we -- we have to keep focused on. Regardless of when -when we send these scores out or not. And I think agree. I do agree. People want to know.

No matter how many people did not participate in these tests, I think, because I've talked to them, families, kids, teachers, outright question, do you want to know how you did, at least? Of course.

UNIDENTIFIED VOICE: Thank you.
MADAM CHAIR: Deb?
MS. SCHEFFEL: So my question. I'm
questioning the underlying assumptions of the way these cut scores are set. So your point, Jane, the people deserve to know something about how they did, yes.

UNIDENTIFIED VOICE: Yes.

MS. SCHEFFEL: But my problem is the underlying assumptions have not been exposed, so that I'm concerned people get these reports and they assume that they're right. My child has low moderate command, and my concern is just let's -- let me just read one -- one example of the performance level descriptors. If you're distinguished, you can analyze the interconnectedness of the world, including how the movement of people, goods and ideas can enrich cultures or create tensions. And how the uneven distribution of resources can lead to conflict, competition or cooperation. So apparently if you can do that, according to what? A rubric, according to a group of people that sat down and said, "This is the way that analysis has to look to be distinguished," and then you're going to distill it into a report like that, that people are going to take away teachers, parents, kids, and say, $I$ need to do $x, y$, and $z$ better or worse.

You know, in other words, I -- I feel the underlying assumptions are off determining how these cut scores are set. So I'm all about talking about a positive result here. How can we give feedback to the field? But if we do it based on this approach, I think there are so many nested assumptions in these cut scores that nobody understands, that it will lead to actions that -- that
people may not have anticipated, and again, don't understand.

MADAM CHAIR: That -- that could be very true, but is it that simply that happens all the time.

MS. SCHEFFEL: Not to this extent. Not -not the way this (inaudible).

MADAM CHAIR: And -- and how -- how would you suggest we reach these scores?

MS. SCHEFFEL: That's what I'm saying. I think that's our next -- that -- that discussion we need to begin. How can we give feedback to the field without doing it this way, with these assumptions that have not been critically analyzed? They've not been socialized with the public, so that they know what those scores mean. That to me is the problem.

MADAM CHAIR: And what would you -- and I'm sorry, but we did have this a good conversation.

UNIDENTIFIED VOICE: You were saying that you have curriculum.

MADAM CHAIR: Excuse me.
UNIDENTIFIED VOICE: Well -- well --
UNIDENTIFIED VOICE: I'm sorry.
MADAM CHAIR: What do we do, Deb, in the, yeah, and I'm (inaudible). What do we do in the meantime, you know, school's going on, teachers are giving
assessments, they're giving grades.
UNIDENTIFIED VOICE: That's why I'm asking, what are our options?

MADAM CHAIR: What do we do about that?
UNIDENTIFIED VOICE: Can we give raw scores, mean scores, ranges? I mean, can we give some other type of feedback that doesn't have so many nested assumptions that have been, have not been socialized with parents or teachers or kids, so that they'll get that report. and they don't know what it means?

MADAM CHAIR: I don't know. I -- I, you know, I -- I --

UNIDENTIFIED VOICE: I think it's possible.
MADAM CHAIR: -- just --
UNIDENTIFIED VOICE: We could (inaudible)
MADAM CHAIR: -- I don't want to harm the ongoing people in the classroom now that are giving these kinds of things.

UNIDENTIFIED VOICE: My concern is --
I don't want to say don't --
UNIDENTIFIED VOICE: -- that we do more harm.

MADAM CHAIR: -- what?
UNIDENTIFIED VOICE: My concern is that we do more harm in releasing this information --

MADAM CHAIR: So --
UNIDENTIFIED VOICE: -- this information.
And the assumptions haven't been critically analyzed.
MADAM CHAIR: And I -- I -- I -- I don't know. I've argued with you, I understand. But what do we do right now? What are we do for the people in the classroom right now? How do they determine whether their students are -- are making the grade or not making the grade?

UNIDENTIFIED VOICE: Well --
MADAM CHAIR: And how do they report to parents, and how do they report the kids if we --

UNIDENTIFIED VOICE: -- that's why we're (inaudible).

MADAM CHAIR: -- don't have any cut scores?
UNIDENTIFIED VOICE: I'm -- I'm asking for us to discuss. That's why we're discussing it. What are our options? What kinds of data could we provide?

MADAM CHAIR: And -- and this might be a long, long, long discussion on giving us time. We don't have time for it here this long discussion. Go ahead Angelika.

MS. SCHROEDER: So I'm going to suggest, Deb, that when teachers get this feedback I don't think they're going to be confused, because we went through a two year
process to develop these standards. And they have been adopting curriculum to in order to implement that. I'm going to totally agree with you that we have got a communication piece to do, because much as this is interesting, it is like way over my head. And I'm familiar with some of this terminology, and it still is really hard. So I completely agree with you that we really need to in our reporting be able to speak to this. For me it has to have some examples.

You know, a rubric for me means something when it actually has a particular issue, and there are examples of what is at which level. So I'm going to agree with you on that part that we've got to do some serious communicating, because the expectations that we have today for kids are different from the expectations that we had experienced as -- as students ourselves.

UNIDENTIFIED VOICE: And the teachers haven't read these -- these --

MS. SCHROEDER: No.
UNIDENTIFIED VOICE: -- performance level
descriptors?
MS. SCHROEDER: Oh, yes, they have.
UNIDENTIFIED VOICE: They understood --
MS. SCHROEDER: Oh, yes they have.
UNIDENTIFIED VOICE: -- what the various
levels are?

MS. SCHROEDER: Yes, they have.
UNIDENTIFIED VOICE: It hasn't been voted on yet.

UNIDENTIFIED VOICE: It's -- there's so much of that in the standards.

UNIDENTIFIED VOICE: It's (inaudible). UNIDENTIFIED VOICE: Madam Chair.

MADAM CHAIR: Yes, ma'am.
UNIDENTIFIED VOICE: So, and again I think when we were here in August, the educator tried to address this. And he articulated that, at least from his perspective, that there was a pretty clear link between those performance level descriptors and the standards, so that those folks who are teaching social studies, they seem to be able to look at that, and say, I understand how this is building. I know how to make sense of this. I agree that is that everywhere at this point in the state? It's not. But I'm not sure it's going to get everywhere in the state until we actually put results out there, and they're actually linked back to those performance level descriptors.

One other thing that $I$ do want to mention too is that our expectation is that we will continue to release items that have --

UNIDENTIFIED VOICE: That will be huge.
UNIDENTIFIED VOICE: -- that have not just the item, and not just the correct answer, but also will have what the performance looked like on that item. So what percentage of kids actually got that right? And when it comes to the constructed response, what did a zero look like compared to what a two look like? How do we start to understand what a two performance is on some of our constructive response items? And again, at this point we've had about 300 Colorado educators who have been involved in this process. There's a whole bunch more, but we have to get the information out there in order for it to be used.

UNIDENTIFIED VOICE: This is just a -- a thought. Is it possible for us to -- because it seems like what you're combining here in these reports is Criterion reference and norm reference, right?

UNIDENTIFIED VOICE: Mm-hmm.
UNIDENTIFIED VOICE: You're doing
comparisons as well as to what extent a student approaches a criteria. Right. So for us to circulate norm referenced type information, wouldn't be incompatible with the system that we have set up.

UNIDENTIFIED VOICE: Madam Chair. So I think I heard --

UNIDENTIFIED VOICE: (Inaudible) scores, mean scores, (inaudible)?

UNIDENTIFIED VOICE: Right.
UNIDENTIFIED VOICE: So when we are looking up here at this table, which is the top table on the reports that you have, those littler diamonds represent the mean score for the school, the mean score for the district, and the mean score for the state. So we were very much coming from that perspective of can we provide some more descriptive statistics that folks might be able to understand. And when we shared this with the parents, and I agree with you that when we talk about parents, there's a wide variety of parents, and there's a wide variety of parents in terms of what kind of information they want to have. And these reports, you can stop here or you can continue to dive more deeply.

But what the parents were able to understand is, I'm going to look at where my child's diamond is. I'm going to look to see which performance level, that's going to give me the gauge in terms of how well we're meeting the expectation of the standards. But then I'm also going to look, and if my child's diamond is to the right, I know my kid did better.

UNIDENTIFIED VOICE: Okay.
UNIDENTIFIED VOICE: If my child's diamond
is to the left, my child didn't do as well as. So I think, and -- and again, I'm not saying that we went as far as you may have liked us to, but we tried to address some of those concerns about, give me some other comparative information on these reports that goes beyond just the standard --

MADAM CHAIR: (Inaudible) because we're not got to get to (inaudible).

UNIDENTIFIED VOICE: -- that we're all
trying to learn.
UNIDENTIFIED VOICE: Madam Chair?
UNIDENTIFIED VOICE: So just a follow up, and then I'll --

UNIDENTIFIED VOICE: Okay.
MADAM CHAIR: Yes.
UNIDENTIFIED VOICE: Well, my only thought would be would it make sense for it as I'm struggling with this to release mean and average scores ranges, so descriptive type information give a close up sense of how they're doing, relatively speaking and not the cut scores while we examine the underlying assumptions.

UNIDENTIFIED VOICE: At this point --
UNIDENTIFIED VOICE: (Inaudible).
UNIDENTIFIED VOICE: -- Madam Chair, that does not meet the state or federal stature.

UNIDENTIFIED VOICE: Yeah. What's wrong

```
with doing it in -- doing both, Deb?
    UNIDENTIFIED VOICE: May I -- may I just say
    MADAM CHAIR: Yes.
    UNIDENTIFIED VOICE: -- Madam Chair. I want
    to get it kind of down to the children. The children, and
    what they're learning, and I, you know, Denver Public
    Schools came out, and said they didn't have material for K
    through three. They -- they don't have material. We have
    to listen to what the also, these experts who came to us
    last week, and said --
    MADAM CHAIR: (Inaudible).
    UNIDENTIFIED VOICE: -- it was an evolving -
- an evolving test.
    UNIDENTIFIED VOICE: Madam Chair?
    UNIDENTIFIED VOICE: -- and, you know, I
mean, I -- it just so evolving that we can't really --
    UNIDENTIFIED VOICE: Did you find it?
    MADAM CHAIR: All right.
    UNIDENTIFIED VOICE: -- really discuss it --
    MADAM CHAIR: We need --
    UNIDENTIFIED VOICE: -- and we're -- we're
doing cut scores when people don't understand.
    MADAM CHAIR: -- we need to wrap this up,
    because we don't have all day. What did you say,
```

UNIDENTIFIED VOICE: Well --
MADAM CHAIR: Oh, you didn't? Okay. I have a -- you -- you have sent -- I will let you speak since you're the only one who hasn't.

UNIDENTIFIED VOICE: (Inaudible).
MADAM CHAIR: (Inaudible)
UNIDENTIFIED VOICE: I have a number of questions. Just kind of basics.

MADAM CHAIR: Well, if you've got a number of questions, let me speak first, because we do have a very limited amount of time. Obviously we're not going to come to a conclusion here. And --

UNIDENTIFIED VOICE: We need to --
MADAM CHAIR: -- I -- I have a motion here that $I$ could use to table this, and I think at, you know, at this late hour of the day with tempers fraying, and, probably not a solution in sight unless we to --

UNIDENTIFIED VOICE: (inaudible) temper (inaudible)

MADAM CHAIR: Would you not do that to me, please?

UNIDENTIFIED VOICE: I'm sorry. Forgive me.
MADAM CHAIR: I think that that it probably would be, I would advise that we table it and come back to
it, because 4:30 in the afternoon, I don't think we're all up to it. But if you want -- if you just make it brief. Okay.

MR. DURHAM: Well, yeah. I do have just a couple of questions. One is where -- currently, who holds the data for the individual students? Is it held in this department? Is it held by the districts who administered the test? Who -- who holds the data?

UNIDENTIFIED VOICE: Madam --
MR. DURHAM: Who produce these reports?
MADAM CHAIR: Yes.
UNIDENTIFIED VOICE: Madam Chair. So once the reports are produced, that data will be at the school district and state level.

MR. DURHAM: Who holds it now? The school districts don't have it?

UNIDENTIFIED VOICE: Madam Chair. The school districts do not have this data now, because we can't fill in this field of what ultimately is the student's performance level. So there's a hole in the data at this point in time.

MR. DURHAM: Okay. So then -- then as I understand it, this data cannot be used for teacher accountability under the statute that just passed?

UNIDENTIFIED VOICE: Correct.

MR. DURHAM: That it can't be used for district accreditation under the standard that just -- just passed. Is that correct? Can you --

UNIDENTIFIED VOICE: I believe it's possibly -- possibly for the first.

UNIDENTIFIED VOICE: You're correct.
MR. DURHAM: Yeah.
MADAM CHAIR: You got it.
UNIDENTIFIED VOICE: And -- and, yes, for the second. I -- I think -- I think it's permissive under -- under the -- the teacher.

MR. DURHAM: That's not -- not state required?

UNIDENTIFIED VOICE: I'm sorry.
MR. DURHAM: And so it -- it is -- it is -it does -- it -- if released would provide some information to individual student on their performance. That's the one thing it would do, and I think the most compelling argument that the Chair made is that students, some of them, some at least may be interested in this data, although given the number who didn't take the test, it may not be as large as we think.

UNIDENTIFIED VOICE: It's 82 percent.
MR. DURHAM: Secondly, the comments that we'd be out of compliance with state and federal law, I'd
just like to make one comment. The legislature cannot compel any member of this Board to cast a vote any way that they don't want to, number one.

If, in fact, this Board decides not to release these scores, there is no penalty then that accrues to any one individual or to the Board for that matter. And if somebody wants the scores released, their option is to go to court, and -- and have the court release them, since we fail -- since we failed to do it, and it gives us a chance to make a legal case about whether their valid or not. So I'm not too worried about -- about that.

But I -- I do think that --
MADAM CHAIR: You're not worried about not releasing the scores?

MR. DURHAM: I -- I don't think there's -- I think that the -- the biggest damage comes from individual students who took the test, and are, in fact, interested in the results.

MADAM CHAIR: Well, yeah, I think that's --
MR. DURHAM: I think -- I think that's, that's the problem.

MADAM CHAIR: You're right.
MR. DURHAM: So, you know, I'd hope we might be able to come to some sort of compromise on this, but the more it -- the more I hear, the more I -- I guess less
likely I am to compromise, because I keep falling back into the mode that -- that these could very well be designed to create a crisis in education for the purposes -- for political purposes with which I don't happen to agree.

So, but taking that out of it, I'm -- I'm certainly -- certainly willing to go along with the motion to table, but I think we ought to have -- I really think we ought to have enough time to discuss whether we could release this for some purposes and not others. That is not for district comparisons from district to district, but were only -- only released on an individual basis to individual students, so that wouldn't affect teacher accountability, wouldn't affect district accreditation.

In fact, the press would not have the opportunity to make those comparisons, because that data wouldn't exist to be able to compare Cherry Creek to Cheyenne Mountain, for example.

And -- and we might be able to do some of those things and minimize -- minimize the press damage, which is going to be that, you know, Colorado students and teachers are all failures, which is going to be the headline when this data is released and we have a crisis. So --

MADAM CHAIR: But we also have the crisis when they take -- the State Board refuse to -- to release
this information, and then there's going to be a second crisis.

MR. DURHAM: Well, I don't -- I don't happen to believe that's crisis, because if -- if the information -- if

MADAM CHAIR: (Inaudible).
MR. DURHAM: -- if -- if the majority of this Board believes that the information is inaccurate in some fashion, that's our position and I think it's a defensible one. And I think -- I think Dr Scheffel made a pretty good prima facie case at least that they're questions about the data.

MADAM CHAIR: Okay.
MR. DURHAM: So can we get --
UNIDENTIFIED VOICE: (Inaudible).
MR. DURHAM: -- a long enough discussion at some point to see if we can work out some compromise based on for what purpose it might be released, in what format it might be released, to whom and might be released? And we can maybe deal with the simplest (inaudible).

One -- one last question is, if we set this as a cut score --

MADAM CHAIR: Well --
MR. DURHAM: -- it's -- is it only for this year, and only for this -- does this carry through in any
way for the future?
UNIDENTIFIED VOICE: Madam Chair?
MADAM CHAIR: Yes.
UNIDENTIFIED VOICE: Cut scores can be evaluated on an ongoing basis. Typically, you don't change them year after year after year, because that eliminates the ability to really evaluate whether or not there are changes within your system. But I would encourage us, especially given the circumstances that we experienced this fall, that we have a clear plan for doing ongoing investigation about whether or not these scores are indeed at the right spot. And I would love to be able to see how students perform without what I'll call all the extra noise.

MR. DURHAM: Well, I'll just make one --
MADAM CHAIR: Go ahead.
MR. DURHAM: -- final closing comment then.
The, you know, I think in all the debate that went on at the Capitol, one thing became very clear, and that the -the -- the group that has no stake in this, at least a stake that they can perceive are the students. And they are opting out -- in the -- in the -- at senior level, they voted with their feet. They don't believe, at least, they have a huge stake in this.

It's quite clear the teachers don't think
they have a huge stake in this, and/or that they do have a huge stake, and that it's improperly applied. Parents I think really don't know at this point in time how to evaluate this, because of $I$ think the noise in the system, which, you know, I -- how we got here is by the way of going to be a fun debate.

But I think we need to set aside enough time to see if we can work through some compromise or for some release of some of the data, but $I$ think a full release, even with these modified levels, I'm still not willing to vote for.

UNIDENTIFIED VOICE: Madam Chair?
MADAM CHAIR: Yes.
UNIDENTIFIED VOICE: Mr. Durham, that's a question maybe you can answer. I don't know how we can -don't know right now, and that's something we have to get back with how do we parcel it out. Okay. That's some way, I don't know if that's possible, but that's something if you want we will get back.

I don't know -- I know, because the system, how it's designed. Okay.

UNIDENTIFIED VOICE: Come on.
UNIDENTIFIED VOICE: Madam Chair?
MADAM CHAIR: Okay.
UNIDENTIFIED VOICE: And respectfully, I'm
willing to look at all kinds of different options, and take into consideration all of your suggestions. There is a little bit of a conflict between what I'm hearing on this side, which is please be sure to give us the comparisons of the student, against school, against district, against state for that comparison purposes. And then saying we won't release it.

I -- I just have to be up front about that. That almost by default, even if we would somehow say we're only releasing the individual student reports, by default there are -- there's school data, and there's district data.

MADAM CHAIR: Okay.
UNIDENTIFIED VOICE: Again, not that we can't make shifts. I'm trying to figure out how to meet both expectations.

MADAM CHAIR: And we need to wrap this up quickly. Angelika.

MS. SCHROEDER: So factual information, 82 percent of the seniors took those two tests, so it's not, yes, 18 percent of the kids in those 82 percent of the kids who did take it are now being cheated because we are not providing information. Their parents are not getting the information.

Nothing is going to change about how, based
on these new hard standards with a different framework than we've ever used before. It's the same framework NAPE's been using, the AP has been using. It's a lie for us to say it's something different than what one of these three is, because this is, these are the results. What we want to see is improvement. We want to see kids who aren't near the top getting more help, but to say this is not the truth, there's just nothing that supports that comment, because this is the truth. It's substantiated by a whole lot of other assessments that our kids have been taking. MR. DURHAM: Well, I'd love to have an hour to debate that with you, because I think you're wrong.

MADAM CHAIR: I would love to add to that. I think perhaps some of you don't have any idea what are -we're not releasing test scores. Kids and parents are not going to know what they did, and because of our actions, and you think that this is not going to raise a huge storm in the state. What do they think they're doing up there?

This is not good -- good for the State
Board, and I would -- when you talk about the number of students who opted out, they opted out with their feet. They were whipped up to that. There -- there was a lot of publicity going on at -- that was not true, was not right, but, you know, they bought it. And kids loved to opt out. I mean, that's the best thing they can do is opt out.

UNIDENTIFIED VOICE: I (inaudible).
MADAM CHAIR: So I think --
UNIDENTIFIED VOICE: (Inaudible).
UNIDENTIFIED VOICE: I wanted to make to
make a motion. Do we have it or not?
MADAM CHAIR: I think if -- I'm ready to offer the motion that we table this, but if we table it , then we're not talking about it till June.

UNIDENTIFIED VOICE: This is ridiculous.
MADAM CHAIR: And the kids are out of school in June. I just think it's -- it's, you know --

UNIDENTIFIED VOICE: It's unprofessional. It's just plain unprofessional.

MADAM CHAIR: Unprofessional, good word. It just boggles my brain that we would do this. But, you know, Angelika, would you perhaps do the motion? Would you say that?

UNIDENTIFIED VOICE: I'd love to make a motion.

UNIDENTIFIED VOICE: Go for it. Make a motion.

UNIDENTIFIED VOICE: Well, I like to make a -- a motion that we give grades a, b, c, and f, so --

MADAM CHAIR: I can't change the whole thing like that.

UNIDENTIFIED VOICE: You can't do that. MADAM CHAIR: You can't, you know. We either release the grades or we don't release the grades. We can't just suddenly come up with a new grade scale.

UNIDENTIFIED VOICE: I don't see why not out of (inaudible).

UNIDENTIFIED VOICE: I mean, this would be probably more meaningful then --

MADAM CHAIR: (Inaudible).
UNIDENTIFIED VOICE: -- it probably would be more meaningful to parents then what they'll be getting with that -- with that score scale or with that scale. I think they will understand $a, b, c$, and $f$. So --

MADAM CHAIR: I think the feelings are hot enough that we're going to wrap this up.

UNIDENTIFIED VOICE: It's not hot enough.
UNIDENTIFIED VOICE: So I move to table the action on the high school science and social studies recommended cut scores, and their associated performance -performance level descriptors for the Colorado measures of academic success, and Colorado alternate -- alternate assessment until the June Board meeting.

MADAM CHAIR: Is there a second? No? No second?

UNIDENTIFIED VOICE: Not me.

MADAM CHAIR: You want to sit -- sit here and continue --

UNIDENTIFIED VOICE: No. I'll make another motion.

UNIDENTIFIED VOICE: There's no second. UNIDENTIFIED VOICE: Well, why not take up my motion.

UNIDENTIFIED VOICE: I'll make a substitute motion.

MADAM CHAIR: That we sit here all night?
MR. DURHAM: I -- I've been with the legislature the last two -- last two weeks.

MADAM CHAIR: You're used to that.
MR. DURHAM: I -- I can promise you I can stay all night. I'll make a substitute motion that -- that this layover until tomorrow, and that we devote at least, find the time to devote 30 minutes to it and see if staff can come up with some compromise that we might be able to live with that incorporates, because it incorporates at least some of the concerns that they've heard. Maybe they can't, but we'll see if they can.

MADAM CHAIR: That's a good motion.
MR. DURHAM: And so that's --
MADAM CHAIR: And I think Jane's going to

```
second it.
```

MS. GOFF: And I will second it, because we just happen to be in the same direction.

MADAM CHAIR: All right. Then moved and sect that we -- that we set aside 30 minutes tomorrow, which will not be easy. We've got a very full schedule tomorrow, but we'll find 30 minutes, and try to come to some conclusion. Maybe (inaudible) have some suggestions for us.

I really think people this is so important.
UNIDENTIFIED VOICE: Stay up all night.
MADAM CHAIR: We are, you know, we're just letting ourselves in for a huge firestorm if we don't release test scores this year. I just -- I -- and we can talk about the fact that we need to do further study on how to norm them, and all that sort of thing, but these tests have been given, and the fact that we won't release them is just huge.

UNIDENTIFIED VOICE: I --
MADAM CHAIR: With that we would move on to the next item on the agenda.

UNIDENTIFIED VOICE: Madam Chair.
MADAM CHAIR: I will (inaudible) -- oh,
(inaudible). Yes. Bizy, will you call the roll?
MS. BURDSALL: Steve Durham?

MR. DURHAM: Aye.
MS. BURDSALL: Dr. Flores?
MS. FLORES: Aye.
MS. BURDSALL: Jane Goff?
MS. GOFF: Aye.
MS. BURDSALL: Marcia Neal?
MADAM CHAIR: Aye.
MS. BURDSALL: Pam Mazanec?
MS. MAZANEC: Aye.
MS. BURDSALL: Dr. Scheffel?
MS. SCHEFFEL: Aye.
MS. BURDSALL: Dr. Schroeder?
MS. SCHROEDER: Aye.
MADAM CHAIR: Thank you. Thank you for the very productive discussion.

U UNIDENTIFIED VOICE: So much, I think I had your book by accident. Sorry.

MADAM CHAIR: That's okay.
(Meeting adjourned)

CERTIFICATE
I, Kimberly C. McCright, Certified Vendor and Notary, do hereby certify that the above-mentioned matter occurred as hereinbefore set out.

I FURTHER CERTIFY THAT the proceedings of such were reported by me or under my supervision, later reduced to typewritten form under my supervision and control and that the foregoing pages are a full, true and correct transcription of the original notes.

IN WITNESS WHEREOF, I have hereunto set my hand and seal this 22nd day of January, 2019.

/s/ Kimberly C. McCright<br>Kimberly C. McCright<br>Certified Vendor and Notary Public

Verbatim Reporting \& Transcription, LLC 1322 Space Park Drive, Suite C165

Houston, Texas 77058
281.724.8600

