COLORADO
Department of Education

TRANSCRIPT OF PROCEEDINGS
BEFORE THE
COLORADO DEPARTMENT OF EDUCATION COMMISSION DENVER, COLORADO

November 12, 2015, Part 2

BE IT REMEMBERED THAT on November 12, 2015, the above-entitled meeting was conducted at the Colorado Department of Education, before the following Board Members:

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

CHAIRMAN DURHAM: All right. We'll now proceed with item -- item 4, which is the -- the results of the PARCC English Language and Math tests, which are being released probably as -- as they become public here. And may I ask, Joyce, how long is your -- roughly how do you tend to go at this, or how do you intend to present this?

MS. ZURKOWSKI: Mr. Chair, we do have obviously the results to share with you. We expect that there will be some conversation that you are going to want to have about the results, and perhaps some other things. I believe we have an hour scheduled for this conversation.

CHAIRMAN DURHAM: Okay. So then you want to proceed then with your basic outline?

MR. ASP: Mr. Chair, could I add just --
CHAIRMAN DURHAM: Yes.
MR. ASP: -- one comment? Thank you.
CHAIRMAN DURHAM: Commissioner.
MR. ASP: Just to remind us all that this is the first release, if we hadn't remembered that, of these results from the assessments for English language arts, as part of our Colorado measures of academic success. These successes were developed by the PARCC Consortium.

You can see from the media attendance, as
well as attendance from school districts, and advocacy groups, and higher ed, that there's great interest in this, and we've been excited to share these results with you since they are the first time we've looked at English language arts assessments that -- that are aligned with the Colorado academic standards. With that, I'll turn it over to MS. Zurkowski.

CHAIRMAN DURHAM: All right. And let me say, we'll -- we'll try and -- we'll hold the questions, I think, until we're through with the initial presentation, and the results are all in front of us, and then we'll proceed with questions.

MS. ZURKOWSKI: Mr. Chair, so in terms of a high-level outline, $I$ want to provide a little bit of a background, in terms of these assessments; talk about the performance levels; and then get right into those participation and achievement results. Alyssa will join me and talk a little bit about accountability. We'll talk about the release schedule for school and district level information, show you some sample score reports, and point to a couple of resources for schools and districts.

So in terms of what was the rationale from -- for moving from CCAP, TCAP, to the CMAS
assessments, as you know, Colorado adopted new standards in December of 2009. Updated those standards in August of 2010. And we needed to have an assessment that actually measured what those standards were.

The rationale for moving forward with new standards and assessments, folks really in Colorado did some deep dives into what was happening with students who were exiting our $\mathrm{K}-12$ system, and they knew that we had about 34 percent of graduates from the class of 2013 who went to public colleges in Colorado. They were in need of remediation in at least one class. We know that Colorado is only producing 22 college graduates to every 100 students who enter a Colorado high school, and so the goal was to increase the readiness of our students for both college and careers.

The Colorado academic standards were fully implemented in 2013-14, so last year was the second year of full implementation. Some districts did move more quickly. So they are in their third or fourth year of implementation. Again, the expectation is that these new and more challenging standards will better prepare our students for college and career.

The CMAS tests in English language arts and math, as well as in science, and social studies, are aligned to those standards. As the interim commissioner
indicated, the Colorado measures of academic success actually consist of two different components. The one is the Colorado developed science and social studies. And then the other side are the PARCC developed English language arts and mathematics assessments. Today, we are talking about those English language arts and mathematics results.

As I indicated earlier, again, these tests are aligned to our Colorado academic standards. They were designed to be administered online. That was a priority for Colorado during their conversations back in 2009 and 2010. They feature more interactive and engaging questions. And they assess concepts and realworld skills that are included in the standards. Just as an example of some of the changes that we saw, and we do see some of the most significant changes within the area of English language arts, what you have on the left is an example of a middle school TCAP reading prompt. And it says, "I am most proud of the fact that this school year I blank." And students were asked to complete that and write to that.

With the new English language arts assessments an example is, you have read a passage from "The Count of Monte Cristo," and a scene from "Blessings." Think about the similarities and
differences and how the two authors developed the themes in each text. Write an essay in which you identify a theme from each text and analyze how each theme is developed. Be sure to include specific details from selections.

Very different types of writing that we're asking the students to engage in. Much more academic. Much more evidence based than what we have had historically.

As we get ready to review these results, just wanted to remind us some of the purposes of our state assessments. They do serve as one indicator of student mastery of the grade level standards by the end of the year.

Couple of points: this is intended to be an end-of-the-year assessment. It is not an interim assessment given throughout the year. It's an end of the year -- did the student, by the end of the year, master those concepts and skills for that grade level. And also, it's one indicator. We expect that schools and districts will take this information, and as they make decisions about decision -- sorry -- about students, they take this into consideration, as well as what they have locally.
how students are performing compared to their school, compared to their district, and to their state peers. And that is relatively unique to the state assessments. With the ELA and math assessments we also have comparison to other states. Again, most local measures don't provide that kind of information.

As we move forward, the assessments will track yearly student growth. They also allow teachers to see how their students are performing against the standard and identify areas they may need to adjust in their practice for the future. So this is always looking at future; what can we do to improve. They also do provide school and district comparisons, and the accountability information for parents, students, and the community.

For the 2015 administration, we administered English language arts to students in grades 3 through 11. We administered math to students in grades three through eight, as well as students in high school. We had two different sets of high school assessments: the traditional pathway of algebra, geometry, algebra II.

There was also a pathway for what we refer to as the integrated or international pathway 1, 2, or 3. And at the high school level students had some flexibility, in terms of which of those assessments they
would take, based on what most closely matched what they were taking within their classroom experiences.

I also wanted to mention that as we made the move to online assessments, we were able to increase some of our accessibility features and what we have had historically. We also had accommodations. When we're talking about accessibility features we're talking about students being able to use things like highlighters. We're talking about students being able to use bookmarks -- electronic bookmarks. We're talking about students being able to eliminate answer choices. They also are able to get directions clarified. There are things, like, word-to-word dictionaries, as well as popup dictionaries.

For students, who are English learners, and are engaged in bilingual education with both English and Spanish, there was for math a Spanish online version of the math assessment that also had what we referred to as text-to-speech, so the student could hear, as well as see that those words in Spanish. There was also a Spanish version of the assessment.

At the local level folks were allowed to translate that math assessment into languages that were most appropriate for their students, based on their instruction within the last year.

With these new tests we are setting a new baseline. And again, the focus is on college and career readiness. They are new tests. As I indicated earlier, the first year of implementation was '13-'14 for those new standards, so last year was the second year of full implementation.

We expect that as students and teachers become more familiar with those expectations, we will see a rise in scores. Please keep in mind that making comparisons between the CMAS scores and TCAP scores, avoid doing that. These are very different tests, assessing very different sets of concepts and skills than what we have had historically. And again, this year's scores will serve as the baseline to measure future student growth, and school, and district improvement. Performance levels. For purposes of PARCC reporting, the performance level cut scores were set this past summer -- thank you -- and early fall. There were approximately 200 educators from across the consortium that participated in that process. We had appropriate 25 Colorado educators who participated in that process.

The PARCC governing board adopted the cut scores for purposes of the PARCC reports. Each state will make its own decisions regarding accountability reporting. And Alyssa will address that in a little
while.

What are the performance levels for PARCC?
There are --
CHAIRMAN DURHAM: Could you just repeat that last thing -- each state will do what? And could you explain it just a little? I'm sorry.

MS. ZURKOWSKI: Each state will make their own decisions regarding how they utilize the results in accountability decisions

CHAIRMAN DURHAM: Thank you. Sorry.
MS. ZURKOWSKI: Yep.
CHAIRMAN DURHAM: Violated my own rule about interrupting, so --

MS. ZURKOWSKI: You're the Chair.
So --
CHAIRMAN DURHAM: Why not?
MS. ZURKOWSKI: -- five different
performance levels for PARCC: the lowest level is level 1, did not yet meet expectations; level 2 is partially met expectations; level 3 is approached expectations; level 4 met expectations; level 5 exceeded expectations. Level 4 and 5 are the levels that correspond to being on track for being ready for the next grade, or for being college and career ready.

We have received questions about how does
this compare to our science and social studies cut scores, or performance levels rather. For science and social studies our distinguished command is comparable to the exceeded expectations; strong command is comparable to our met expectations. Again, the top two levels are the indicators of being on track, whether that be for the next grade, or for college and career readiness.

Remember, that we had moved forward with science and social studies prior to ELA and math, so in the process of the last two years additional decisions have been made by the PARCC states. We will be making some adjustments to how we refer to our performance levels in the future, so that we aren't asking our field to have to balance two different sets of vocabulary. So for 2016 again, we will bring those more into alignment. And again, you can see how those correspond moving down. UNIDENTIFIED VOICE: May I ask a question? CHAIRMAN DURHAM: Let's go ahead and finish --

UNIDENTIFIED VOICE: Oh, sorry.
CHAIRMAN DURHAM: -- if we can.
MS. ZURKOWSKI: Before we get to the results, $I$ wanted to talk a little bit about participation. I think that's been a -- I'm going to give that to you -- that has been obviously a topic of a
lot of conversation. You can see what I provided here is participation by grade level, and by content area. There's not a lot of difference between the content area, right. So if we had a student participate for ELA, we tended to have that student participate for math.

When we look at our lower grade levels, grades three, four, and five, we are very close to that 95 percent target. When we look at grades six, seven, and eight, there is a little bit of slipping, and it's between about 85 percent and 93 percent. When we get to high school, we see a more significant drop in participation. Ninth grade at about 70 percent. Tenth grade at about 60 percent. And 11th grade at about 50 percent.

We do not provide grade level participation for math at grade 11, because students were not required to take a math assessment in grade 11, depending upon where they were within their coursework flow. So we don't really have a denominator for that particular assessment.

Keep in mind that in the legislation that was passed last spring that PARCC will not be given to students in grade 10 and 11 in this upcoming year.

As we look at who participated, it's important to also look at our subgroup information,
right. It's one thing to say that 95 percent of our students participated, or 80 percent of our students participated, but we should ask which 80 percent participated, right. What -- what does that group look like?

So what you have in the next set of slides are a breakout of who participated in the assessment by subgroups: gender, race, ethnicity, free and reduced lunch eligibility, disability status, and English learner status. In the second column is what we would expect the split to be, based on enrollment information. And what you have in the last column is what the actual participants looked like.

So we expected for there to be about 49 percent of our testers to be female and 51 percent to be male. When you look at grades three through five in ELA, we landed just about right there. When we look at the breakout for our race ethnicity, again, there's a very close match between who actually tested, and who we would have expected to have tested had all kids participated in the assessment. Free and reduced lunch, also very close to being on target. Disability status, again, very close. English learner status, again, very close for grades three through five.

When we look at grades six through eight in

ELA, we see a very similar pattern, meaning our students who tested represent our overall student population very well.

When we start looking at high school, we need to be more cautious in our interpretations, because now we do actually see that our students who tested do not represent as well our overall student population. So when we look at, as an example, race and ethnicity, and we look at the percent of Hispanic students we had participate, we expected there to be about 31 percent, in reality there was about 37 percent. When we look at the percent of white students, we were expecting about 55 1/2 percent, there was 50 percent. So we actually had in our testers a higher proportion of Hispanic students and a lower proportion of white students.

When we look at free and reduced lunch eligibility status, we see a higher proportion of students, who are eligible for free and reduced lunch actually participate in the assessment, compared to what we have in the population overall.

When we look our students with disability, that is more representative.

When we look at our English learners, again, that tends to be closer to what we would have expected to see.

The next three slides have the same information broken out for math. Very similar to what we have seen for ELA, so I'm going to ask that we jump to slide number 25, which will be the PARCC results overall for English language arts.

We presented the results broken out by level 1, level 2, level 3, level 4, and level 5. Again, level 4 and level 5 are indicators that students are on track to being ready for the next grade level, or for college, and career. The range is between 37 and 42 percent of our students score at that level 4 or level 5, depending on the grade level.

When we look and ask ourselves is that reasonable, one of the outside sources that we can look at are our NAEP results. And we've talked about that for years, right, is we said how do we get ready for these new results; what can we be looking at to give us a hint as to where they might be. NAEP was one of the recommendations that we said take a look at NAEP. That's probably a better indicator than what CCAP and TCAP were providing for us.

NAEP for grade four and eight we were at about 38 to 39 percent. So our ELA results are consistent with what we see in NAEP.

When we look ACT, we have about 36 1/2
percent of our students who achieve the college readiness benchmark on reading compared to what we had in grade 11 here at about 40 percent. Again, relatively comparable that can allow us to have some confidence in these results, right. We have some outside sources that are consistent with what we are seeing here.

For math -- for math we have this broken out by grades three through eight. We also have each of the high school assessments. It's important to keep in mind that students could start those high school assessments as early as seventh grade. All right. So again, depending upon what their instruction looked like, students could take, as an example, either the seventh grade assessment, or an algebra test.

When we look at eighth grade, students could take the eighth grade math test, or an algebra I test, or a geometry test, or an integrated $I$ test, or an integrated II test. You'll see that when we look at grade eight, that is our lowest percentage of students who are achieving that readiness mark for being on track for the next graded level.

The fact that we had about 10,000 of our eighth graders take algebra, needs to be taken into consideration. So our eighth graders who took algebra are represented in the algebra results. All right. So
our eighth grade results are only for those students who took the eighth grade test. And the range there is between 19 and 37 percent.

When we look at NAEP, again, just as a marker, depending on the grade level, we had between 37 and 42 percent of our students achieve benchmark on NAEP. So again, in terms of reasonableness, relatively consistent.

Breakouts -- first breakout we'll look at is gender, in terms of English language arts. Our females did outperform our males. This is consistent with what we have seen in the past, especially in relationship to writing; that our females write, or achieve higher in writing than our males do.

For math we see a continuation of the trend that we've been seeing over the last few years, which is, our females are catching up to our males, in terms of math achievement, and in some cases are exceeding our males in math achievement.

Next group we're going to look at are -- is by race and ethnicity. We provided information for you for Asians, black, Hispanic, white, and the two, or more race, or ethnicity categories. We did not include some of the other categories, because the -- the number of students we have are really, really low, right. So we
don't include Hawaiian or Pacific Islander. There's just not a lot of those students who fall into that category in the State of Colorado.

What we see is our students belonging to the Asian subgroup are outperforming the other subgroups, including our white subgroup. What we see also is that our black and Hispanic subgroups are relatively similar, and significantly lower than what we see with our Asian, white, and two or more races subgroups.

Historically, our Asian subgroup has outperformed our white subgroup in writing, so there's a continuation of that trend here. It is a reverse for what we used to see with reading. So again, as we're looking at the PARCC assessments, you can see the heavy emphasis in terms of the ability to write well.

For math again, what we see is a very similar pattern with our Asian subgroup performing at the top, then our white subgroup, then two or more race categories, and then with our black and Hispanic subgroups significantly below that. In the case of math, we do have our Hispanic subgroup performing slightly better than our black subgroup. Again, this continues a pattern that we have seen historically.

Free and reduced lunch status. Our students are eligible for free and reduced lunch are not
performing as well as our students who are not eligible for both ELA, as well as for mathematics. Again, this is a trend that we have seen in the past.

Students with disabilities. Our students with disabilities were scoring at a level 4 or 5 between about 5 and 10 percent, compared to our students without disabilities, who were scoring between about 42 and 45 percent on ELA. When we look at math, we see a very similar pattern with our students with disabilities scoring at a level 4 or 5 at a percentage rate of about 5 to 10 percent, and our students without disabilities between about 22 and 40 percent.

For our students who are English learners, we've presented this information broken out by our students who are not English proficient; by our students who are limited English proficient; and then by students who are fluent English proficient compared to our students who don't belong to those categories. What we see is that our not English proficient students are performing significantly below the rest of the groups, in terms of their achievement in English language arts, and that probably makes some sense, right. If you have students who have not yet become proficient in English, they're going to struggle with an English language arts test.

Again, we are not making any kind of determination about our English learners in terms of their literacy in Mandarin, or in Japanese. We're just talking here English language arts. So that's -- that's what we're measuring here.

As students become more proficient in English, you see their scores increase, and in some cases, as we look at our fluent English proficient students, they not only match our non-English learners, but at grade three they are outperforming our English learners. So again, message there being, once our students become English proficient, they are performing at about the same level as their native English speaking peers, or outperforming them.

When we look at math, we see a very similar pattern. With our students who are not English proficient scoring at the lowest end, and then our nonEnglish learners scoring at the highest with a couple of exceptions with our fluent English proficient speakers scoring highest in grade three. Again, these are students who were English learners. They are no longer identified as English learners. They're in their first two years of monitoring, and -- and they're doing very well. So as we're looking at trying to evaluate program effectiveness, folks may want to look at that.

Then moving on to accountability.
MS. PEARSON: Good morning. So we just wanted to talk briefly with you all today. We know assessment and accountability are very closely tied together, and so we just wanted to give a little bit of information about the accountability implications around these results.

First, just a reminder. H.B. 15-1323 created a hold for accountability this year. So these new results that you're seeing are not going into 2015 school and district performance ratings. We have time to help educators understand what the new results mean; get used to them; understand where we're at; have that new baseline set. So there's no accountability based on these results this year.

Right now, based on state law, accountability is set to resume next fall; however, in 1323 it has asked the Department of Education to provide a recommendation to the Joint Education Committees during the smart act hearing, which is scheduled for December 14th, about whether or not our state assessment data is ready for accountability to resume. We will be bringing -- we're working on that as we've -- we're starting to dig into the new data. We'll be bringing that information to you all the December meeting, and
having a conversation with you then before the smart act hearing.

We also know that a large conversation was had around all of the accountability implications with this has been around the participation rates, and around the parent refusals. As Joyce mentioned, looking at this participation rates, especially at the higher grades, they're really critical when you want to go to interpret the scores. So we've been talking about that a lot.

As you all know, you made a motion last February that parent refusals districts cannot be held liable for them during -- in the accountability systems. So we're looking at that data and how to make that work. The federal requirement of the 95 percent participation rate, including parent refusals as non-participant still stands, but with you all approving the ESEA waiver submission yesterday, $I$ think we have a agreement worked out with the U.S. Department of Ed about how to take that into consideration without a -- a level of liability that you were concerned about for schools and districts.

We wanted to share -- go a little bit deeper with some of the participation rate data, so you could see it. This is state wide. We'll see some different things when we look at individual schools and districts, but state wide you'll see here -- this is by grade for

English language arts. The total number of records that -- of records of students that we expected to test, the participation rate overall, and then the number of parent refusals by grade. We know that that was something very interesting, and wanted to show that data to you. And then the rate of parent refusal -- the percent there.

Parent refusals and the data that we have released in participation rates are considered nonparticipants, because we don't have scores for them. So when we're interpreting the achievement data, it's important to realize that they are not part of that -- that data we're looking at.

You'll see in those percentages with parent refusals compared to the participation rate that not all non-participants are parent refusals. There's other reasons why students were -- weren't participating, other than being coded as parent refusals. You'll also notice the percentage of parent refusals increases as you go up in the grades. That's English language arts. And you'll see a very similar pattern for math, like we've seen across the board.

In terms of next steps, like Joyce mentioned earlier, decisions about how these results will be used in accountability come to you all. So the -- per state
statute, the State Board is required to set ambitious, yet obtainable state-wide targets every year. We're thinking in February we'll be ready to look at this data, and will have looked at the data enough, and be able to bring you some recommendations on how we used PARCC results in accountability decisions resuming next fall, and have you all start on that process of setting those targets of what those state level expectations are.

So there's a few different options that we can talk about, about different ways to use the data for accountability, and we'll bring that to you, and have those conversations later.

MS. ZURKOWSKI: Release schedule for school and district level results. Before I talk about school and district, let's about state. So the state results obviously are embargo lifted when we started this presentation. We provided information in terms of achievement by our overall group. We also broke that out by subgroup for gender, race, ethnicity, economic status, English learner status, and disability status.

We have provided information, again, at the state level for participation, in terms of the number of participants, the number of non-participants, including the parent refusals. And we have provided that participation information also broken up by subgroup.

For school and district level results, districts will receive their embargoed district and school results by November 30th. On December 4th we will make information available to the media under an embargoed status, as well as to the Board. Sharing the district and school level results, we expect that embargo to lift December 11th.

Sample score reports are available on the CDE website. Those links are provided here. There is an example of a math high school report, as well as an English language arts grade six report.

This is the first page of the report. I'm assuming that font size is a little bit small to see, so we'll jump in. At the top of the report there will be an indication, in terms of overall performance level. Students will see their skill score between 650 and 850. That will be marked on the color band and parents, and students will be able to see where their score falls, whether that be a level 1, a level 2, a level 3, a level 4, or a level 5. Keep in mind again, level 4 and level 5 are indicators that the student is on track to being ready for the next grade level, or for college, and career.

For comparison information, the school average is provided. The district average is provided.

A state average is provided. And an overall PARCC average is also provided for parents and students to see. For both English language arts and mathematics there are what we refer to as subclaims. Subclaim performance is indicated through arrows. An up arrow means that the student basically met or exceeded expectations for that area. An arrow pointing down indicates that the student is not where we would expect them to be in order to be ready for the next grade level. And an arrow that goes back and forth, side-to-side indicates that they're there. They're nearly where we would expect them to be in order to be ready for the next grade level.

For English language arts there is a reading and writing scale score. The scale for reading goes from 10 to 90. The writing goes from 10 to 60. Information that is provided for comparison purposes is: the school average, the district average, the state average, as well as for those students who overall scored as a student who was meeting expectations what their score on these subscales would be. So for reading where target should be a score of 50 or above. For writing a score of 35 or above. And then below you see the subclaims for English language arts.

Resources. We have put together -- and I
should say that Dana (ph) and her team have put together, I want to make sure we give credit here -- some resources for districts and schools to use as they get ready to communicate with their parents the results.

There's information, in terms of how to help parents understand their students' scores. There's a parents' guide that is available in both English and in Spanish. There is also some information about how to use the test results to support your student. There is also what I'll refer to as an explanation guide to the actual score report itself, so they go through identifying different areas of the report, and then telling parents what each of those pieces mean.

So number 1, that is dealing with the student performance overview. Number 2, is the score range. Number 3, is the indicator of on track for being college and career ready, or the next grade level. Number 4, is where you can find the comparison information. And that goes on.

That is also provided through a Prezi presentation that schools and districts can use, like on a parent's night. There are also some key messages for schools and districts to use, as they develop their messages for their communities, as well as drop-in articles that they can use.

CHAIRMAN DURHAM: Are you -- is that the conclusion of the -- looks like the conclusion?

MS. ZURKOWSKI: That's the end of the slide presentation.

CHAIRMAN DURHAM: Got right on then, didn't we? We're going to do things a little bit different, since I think this is probably the most important issue that we'll deal with certainly in the last six months. And so we're going to do the questioning, and -- and comments a little different. And we'll just start with Dr. Scheffel, and we'll work around, Dr. Flores will be next. Ask all the questions, or make comments, request additional information that you like. Take as much time as you think is appropriate.

We will -- if we run a little bit over on this, I'm not going to get too concerned, so we'll just work our way around the room then starting at this end, and if you think of questions you missed in the first go round, we can come back. So --

MS. SCHEFFEL: A round robin.
CHAIRMAN DURHAM: -- Dr. Scheffel. Yeah.
MS. SCHEFFEL: All right. Thank you.
So it always helps me to think contextually. So when we think of $\$ 360$ million in federal funds with these two testing consortia, one of which is PARCC, many
states have been stepping away from PARCC, and the other the consortia also. Choosing to use their own tests. We're not there today. We're looking at these PARCC data.

So when we look at high stakes assessment, the reason it was put in place is for having a return on investment and metric for public education, given a huge boost with no child left behind, and then continued multiplicity more and more tests. The legislature required us to use PARCC for ELA and math accountability, right. So our task today and going forward is to think about what does the data mean, and as you posed, is it usable for accountability?

And the accountability that it's used for right now is primarily accreditation; is that right? Is there any other accountability relevance, besides the accreditation ratings that are given to districts, based on these data and --

MS. ZURKOWSKI: So -- so school and district -- our district accreditation and school plan type assignments. The school districts may use it for their own accreditation of schools. Right now we're in -- these results can't be used for educator effectiveness for that teacher level accountability, or educator level accountability. In the future I think
it's intended.

MS. SCHEFFEL: But they will in the future. And that's why --

MS. ZURKOWSKI: (Indiscernible).
MS. SCHEFFEL: -- the question are they usable for accountability is so central, right?

MS. ZURKOWSKI: Yeah.
MS. SCHEFFEL: For what? Well, for teacher performance, for -- right now just for accreditation though, right? Okay. And that effects, of course, real estate values, public perception, all of that.

So just, by way of context, I think that helps why are we here today? Well, we're looking at these data. What do the data mean? How can they be used for accountability; in what ways?

So as I look at it, the most important slides to me are 35 and 36, looking at how heavily language loads on these tests. And some of us on the Board took the opportunity to review -- Joyce, you remember that -- the questions on the --

MS. RANKIN: Yeah.
MS. SCHEFFEL: -- questions on the test, both in math and English language arts, and really had a chance to read the items, the stems for the questions, the actual excerpts. And what astounded me walking out
of that opportunity was the heavy linguistic load on the test. And it shows up in very bright lines on these two slides.

So are we testing -- we're trying to get a return on investment. Content knowledge. The ability to apply knowledge and skills. And I really don't believe the tests test that. I think it assesses linguistic acumen. And that's why we see such huge gaps with English language learners until they're proficient in language. So if that's what we intend to test, because we're trying to make meaning of the data, and trying to figure out if it can be used for accountability.

If we're trying to focus on language acumen, then perhaps the test is helpful, but in terms of readiness for whatever we're looking at, workforce readiness, however, we define that; content knowledge. I felt that the test really did not deliver, in terms of content knowledge when I had reviewed it.

So I -- I think that the language piece is so central to the way we interpret these data. I also think that its beyond our -- our purview, as far as authority, to say this high stakes assessment is not helpful to us, but when we see it so heavily loading on language and not on content knowledge, per se, or application. We have to wonder how it should be used for
high stakes decision making. And I think that, to me, is kind of the center piece of our discussion today.

So I think it dramatically disadvantages students who don't -- who -- whose second language is English, and even students that just have not developed great vocabulary skills, or syntactic abilities, or whatever, and so I think it really links to, not so much what students are learning, but their language acumen. So that would be my first comment on these data.

End.
CHAIRMAN DURHAM: Any additional comments in this round?

MS. SCHEFFEL: Not right now.
CHAIRMAN DURHAM: Okay. Dr. Flores.
MS. FLORES: Well, I don't think it -- it -- it really helps poor and minority students. I think that we were doing greater strides when we were really putting ESSA monies to use at the very beginning, and that was during the '60s and the '70s. And we did see that during those years we did have kids -- there was almost parity between minority kids getting to college, and white kids getting to college. And -- and that's kind of really sad for me to -- to think that in a sense I was privileged, and my friends were privileged, you know, because of my age, and because of, you know,
when -- when the baby boomers, I think, went to college. I don't think high stakes testing helps. If you made a comparison between PARCC and NAEP, and you said that NAEP -- there was almost parity, so in essence we have a test, which I think is far better than PARCC, and this is the National Assessments for Educational Progress. And it has been around for a long time. I think the questions may be fair, and it really looks at what people need to succeed.

We know that PARCC is not being used by private schools. It's not being used as in -- to -- in NAEP -- to better NAEP at all. So I think we're using a far -- an instrument that is not as good as, say NAEP. If you want to know how the world is doing, and how each state is doing, we already have an assessment that does that.

Again, I'll say that high stakes testing have never helped kids. We're back to it again. I think we need to get back to the real issues of education, and education actually is about teaching and learning. It is not about accountability. It is not about high stakes testing. And we are spending, and have spent such a large amount of money that $I$ think have -- has been wasted. I think Common Core is not going to help us. It's not going to help us get to better -- better
students. In fact, I think that we have a curricula that has very much narrowed. And it -- our kids are not going to be creative.

We're back to where the Japanese, and the other countries were looking at us to see what was so special about us that we were so creative. And -- and the things that we -- we did, because we were much more open to learning about -- there was much greater knowledge, $I$ guess, that can be found before than can be found in PARCC.

So we should really look at helping our poor and minority kids get better. And I think the way to do that is by focusing on teaching and learning, and we were doing a great job of that $I$ think during the '90s, but somehow we -- we have fallen, and missed our mark, and we're not on focus.

CHAIRMAN DURHAM: Follow up --
MS. FLORES: And these tests show that.
CHAIRMAN DURHAM: -- well -- well, let's -- let's go ahead and go around, then we'll come back. So just make a note of questions.

Dr. Schroeder.
MS. SCHROEDER: So I guess I'd like to get back to the assessment. And I appreciate Dr. Scheffel your comments that we've always talked about the fact
that if kids don't understand the question, they're unlikely -- they're highly unlikely to get the correct response. And this has always been a worry in math, which is the area that -- I'm sorry -- but I feel is just absolutely critical. We know from scores, but we don't know from scores for sure, whether that's a significant weakness in our education system.

So I do feel that this is the first year of this assessment. There's another reason for this time out and maybe additional time outs, which is that we need to have an opportunity to look at the assessment, and to have experts look at the assessment, and evaluate Dr. Scheffel's deep concern about the level of language, and whether that is extremely problematic. I -- I respect that. And I know we had some experts who came -- some psychometricians, if that's the right word, who said sometimes it takes four to five years to really get an assessment the way it ought to be. And so I think we should keep that in mind before we turn ourselves inside out about these results.

I would like to ask the question you -- thankfully you did look at the NAEP, and you saw that we were, I guess, a little higher, even than our NAEP scores, but was the gap similar, because one of the things we're often identified by in Colorado is
having -- is -- is as having some of the largest gaps. Did you see that as well, or did you have -- I'm sorry did you have time to even do that?

MS. ZURKOWSKI: Mr. Chair?
CHAIRMAN DURHAM: Please proceed. MS. ZURKOWSKI: So when we're looking strictly at Colorado data, it is fair to say that the patterns are very similar to what we have seen historically.

MS. SCHROEDER: And it is aggregations as well?

MS. ZURKOWSKI: And it is aggregations as well.

MS. SCHROEDER: Okay. So to look at Dr. Scheffel's question about how should this information be used, I would say number 1, very carefully. Not only because it's the first year, but because we still have things to learn.

Number 2, these kinds of assessments are simply a proxy. We want to know how are the kids doing. And we know that certainly these assessments don't measure everything that goes on; everything that teachers do for kids, so that's why we're talking about our accountability system, and making some changes to it -- changes to it so that we reflect more about how
well our kids doing; what are their environments, et cetera. So I think we -- I think to answer your question, we should just do it very carefully.

It doesn't mean that they're to be thrown out, but it certainly, in my opinion, needs to be about are modifications necessary; what do we learn; what didn't we learn from these assessments. I think that's almost as important as what we -- what we do learn. And then look forward.

I do appreciate very much the very last part of your presentation about how to communicate with families, but also how can families help their kids. When we had -- what was it the Iowa test, or whatever it was that my kids took -- I never had a -- I mean, I got results, but it told me almost nothing, except that they were X number of kids that were stronger students, X numbers of kids that were weaker students, but there was absolutely nothing in there about what her -- their father and I could do to help them; what are the questions we should be asking in the classroom, et cetera. So I am very grateful for that piece of this assessment, because I think it will really resonate in our communities. Thanks.

UNIDENTIFIED VOICE: And most of those tests, could have -- if you had paid more money, you
could have had that data.
CHAIRMAN DURHAM: MS. Goff.
MS. GOFF: Thank you.
I'll start at the end. I -- I wanted to echo appreciation for the communications tools currently developed, and always in progress I hope -- process. I think that's going to be key.

The other -- the other challenge may be for us -- always something to work on is important, I believe I have had more questions regarding the assessment system around the idea why did it take so long for these results, and unless -- unless you are an individual like we are, who live this life, you understand that, or have some background in doing the development, it's very hard for people to understand, especially when there was a lot of touting early on, which we supposedly and -- and did support, for computerized exams -- online exams so that one of the main benefits would be great quick results that are accurate, but I -- I think maybe we've gotten past it now. Maybe it won't be an issue at this point, but from now on -- maybe with the next iteration of whatever we choose to do, that there is some easy to understand language and explanation story around why it takes so long, and the fact that there were educators literally involved in this from the beginning. That's
been a little bit misunderstood.
As far as the exam and results itself, thank you -- thank you both. Appreciate what goes into this, telling the story, and putting together a -- a slide presentation. Along with the language acumen, and -- and those issues, I -- I completely agree it's one of the hardest things to measure is what -- what is the question teaching. And -- and does the question even have a chance of -- of being an -- an instructive took in itself. And then as we -- as we work through that, I think we'll come to some better -- some better outcomes for that.

One thing that was struck from the first time I looked at any of this was well, it sure looks like it, and then hearing, and having reinforced again that these results are -- really do mirror our past, and I kept hearing Joyce's repeatedly -- she said fine -- as former trends have been, or as we have seen, as a trend in the past, and so the question would be, does it -- does the test itself -- the actual product matter. I think the -- the essential question is, why are we seeing the same trends no matter the tool we're using? And this includes both the summative type or the -- the knowledge and skill content check, and obviously the performance check. So we -- we have also had a message
for seven, eight years now in the state. We would like to have demonstrations of how -- what kids have learned, and that they're able to show how that applies to real life, and is that not an ongoing mantra for all of us. I think it is.

So how do we -- how do we get around really just putting ourselves in the box literally of a test that we have to answer the essential question? So to me, that comes back to instruction, strategies. Are we moving with the needs of learners this -- in this day in age? Are we -- are we -- the -- what are we doing to really find out if what we're teaching, and how we're teaching it, and how it's being received is working?

So I think -- I think that's where I am. And then one last point. Couldn't help but notice lower grades, if you -- if you go by the scores, the lower grades are at a higher level right now. Understandable. But the -- the fact that goes along with that are these are the kids -- these are the children who have been living the new standards, and the new ways, and the new strategies, and techniques. They've lived it. Our older grade students have not had that experience, good or bad. They haven't had that experience fully.

And the other outstanding, to me, was I -- I've kind of zeroed in on the title one results
because sort of as fruit from our conversation the last few days over the waiver request, and you know, what do we feel is -- is really it should be a priority for targeting -- targeting money, targeting talent, targeting no how, and expertise. So I'm -- I'm -- for me it -- it gave me a little bit of enlightenment to some conversation about within our -- hopefully a flexibility with our own accountability coming up, and our own ways of looking at resource -- resource allocation and use of dollars frankly.

I think it -- I think that gives us maybe -- does me -- it gives me a little bit more of a pinpointed way to go in the conversation. I appreciate this. I think we've known this wasn't going to be banner all over down day, but on the other hand, it's what it is. So what -- what do we do now? What do we do with this?

CHAIRMAN DURHAM: MS. Rankin.
MS. RANKIN: I think we have a lot of opinions out here. And I think you haven't even heard this beginning of the opinions. So I'll give my opinion. I -- I agree with Dr. Scheffel about the underlying intent content verbiage, whether it's in language arts or math. There's a lot of reading involved, a lot of understanding.

I also look at the type of passages that are put forth to the students and boy, I -- I think -- I don't think you can always get it right the first time around, but I think there's some subjectivity, and I think there's a lot of discussion that's going to come out about this. And I think that's good. I think that's good. I think with challenges come opportunities. We don't even know where they are yet. I mean, we just have the base information. And regardless, the numbers are the numbers. I mean, we -- we -- we can say whatever we want, but it's right there, and I think in February we'll have more information when some of this is digested a little more by the people that do that, but I appreciate how you presented that to us today, and the numbers, and I'm curious to see where this goes. But thank you very much.

CHAIRMAN DURHAM: MS. Mazanec.
MS. MAZANEC: Thank you for the presentation. I also want to echo Dr. Scheffel's comments. (Indiscernible) math tests, but I did go through one of the English language arts tests, and they are very heavily (indiscernible) and I'm concerned about that. (Indiscernible) grammar, and spelling, (indiscernible).

That was brilliant too, and it was --

CHAIRMAN DURHAM: Critical thinking.
MS. MAZANEC: Anyway, so I'm -- I'm -- I
echo her concerns about how language dents all of the testing is, and I again, am concerned about what -- what that's really telling us about how Colorado students are doing. And once again, we have test results that are supposed to tell us how well kids in Colorado are doing, but we're cautioned not to place too much emphasis on them, and we're also told that it'll take four to five years to get these assessments right. And once again, our children are in school now, and being taught to perform on this test. And it seems that every four to five years -- ten years we have another test that's going to be better.

So I understand that we need to put these -- we need to -- we need to consider them cautiously. I just wonder what it means for actual learning for our Colorado students.

CHAIRMAN DURHAM: Let's go back to start with Dr. Scheffel.

MS. SCHEFFEL: Thank you.
I had a question, Joyce, about the proficiency levels. We know that standardized tests have historically been used as measures of how students compare with each other, so norm reference tests. And
then also they can be used to determine how much of a particular curriculum an individual has learned, so that's the criterion referenced approach. And now increasingly we're having standardized assessments used for high stakes assessment, so that goes back to our question of how are we going to use these data.

And as you pointed out, very cautiously, but as -- as we go back to the proficiency levels, and bands, they're subjective, and they're unlike rank orders in -- in a sense, because you can create the cut point wherever you want to. We know that like, I think, in the CCAP, or TCAP the proficiency level for reading, I think, was the 28th percentile. So that was very different information when you think about how well does someone rank, 28th percentile is not very good, and yet, that was dubbed proficient in reading.

So when we look at the proficiency levels on this test, they're subjective. Can you speak to how they are set, and what are the percentile ranks that would be equivalent to those bands, or levels?

MS. ZURKOWSKI: Mr. Chair?
CHAIRMAN DURHAM: Yes.
MS. ZURKOWSKI: So similar to how we've had our conversations in the past about science and social studies, English language arts, and mathematics standard
setting approach that cut score setting approach is very similar. Started off with what are the expectations for students at each level. That's the starting point. So you identify the concepts and skills that you expect from students who are ready for the next grade level, or college, and career ready. I'm using that -- I'm using level 4 as the marker for this conversation.

You are right, in terms of saying that that is based on judgment. That is educated judgment. And I would suggest that we rely on educated judgment a lot, but yes, that did require a lot of conversations across the states; involved K-12 educators; involved higher education folks; was open for comment from Boards, such as yourselves, as well as from parents, and community folks, but what is it that you expect. What are those concepts and skills that you need kids to be able to demonstrate in order to be successful for whatever comes next?

Those descriptions are what guide that cut score setting process. So folks then look at the test and say okay at what point are students demonstrating those concepts and skills. And again, does that depend on educator judgment? It does.

I think I have heard folks here suggest that you rely on educators. You believe that educators are
experts. They were at the table from beginning to end throughout this process. They were at the table from beginning to end, in terms of this process writing the performance level descriptors, participating in item writing, participating in data review, participating in the cut score process. They were the ones who made those decisions.

It was not psychometricians sitting in a back room who made the decisions about the cut scores. Those were the educators who made those decisions.

In terms of course on -- corresponding percentile ranks, I honestly can't tell you that today. I don't have that information, but can we get that information, based on Colorado information? We can. And we can get that to you. Today we don't have that --

MS. SCHEFFEL: (Indiscernible) --
MS. ZURKOWSKI: -- that wasn't part of the process, in terms of what is the corresponding percentile rank.

MS. SCHEFFEL: I think that information --
MS. ZURKOWSKI: Again, it goes back to the content. Sorry.

MS. SCHEFFEL: -- I think that information would be helpful to the public just because it -- it's a traditional way of looking at standardized test results,
and because the bands are so subjective, as -- as indicated with my previous example. I mean, no parent would want to say $I$ feel good that my child is proficient in reading, and then underneath behind the curtain it's, but they're at the 28th percentile. Most parents wouldn't be very happy with that. So I think that looking at both those metrics are helpful when we try to make meaning of these data.

Can you respond to this quote somebody sent me? "The goal of most tests is to sort and rank. To do that, test makers make small differences appear large. Questions most people get right or wrong are removed because they do not help with ranking." Can you respond to that?

MS. ZURKOWSKI: Mr. Chair? So again, when you're talking about what you referenced earlier, normreferenced test, you bet, that is all about ranking. When I am concerned about whether or not I'm the 14th percentile or the 17 th percentile, $I$ am worried about that ranking kind of a concept.

With these tests we're not as concerned about ranking. What we're concerned about is being able to answer the question of what is it that you know, in terms of the concepts and skills that are taught at your grade level. So it is much more content based, and
that's the emphasis, as opposed to can I get each one of you ranked in terms of your knowledge of math. Instead, what I want to understand is what do you know about fractions; what do you know about fractions; what do you know about fractions. And so the test is built more from that perspective, as opposed to a kind of ranking kind of a system.

Historically, I would suggest that with state systems there was a lot of focus on that -- what I'll call that proficiency bar, right, and whether or not kids were above that bar, or below that bar. And there was a lot of focus on making sure that the test did a really, really good job about that bar.

One of the mandates that was given to us was move beyond the bar. We want to know how our kids are doing across the spectrum, so make sure that when kids are getting results, we have meaningful information, regardless of whether they're above the bar, below the bar, but we want to know how they're doing across the spectrum, and so that's really how this test was built, and will continue to be improved. Is to make sure that we are not hyper focused just on this; we are focused on giving meaningful information to all kids.

MS. SCHEFFEL: And so we're back to really to the blueprint of the test, right?

MS. ZURKOWSKI: Yes.
MS. SCHEFFEL: We know that given the proficiency levels, which are subjective, and how the cut scores are set, and what's -- what counts for proficient, or some other level, really deciding what items to include on the test, how the questions are worded, how long the questions are, which answer -- answers are scored correct, how the test is administered, of course, how the exam results are used --

MS. ZURKOWSKI: Sure.
MS. SCHEFFEL: -- and all of that is subjective. And it's based on the blueprint of the test, so as we try to make meaning of the results, we have to look very carefully at what -- what the answers to those questions are, and -- and help the public understand, because this is a narrative of failure in many ways. At least, for some very large subgroups of population, and -- and other groups. So I think that -- that blueprint piece, and how those subjective decisions were made are really crucial as we interpret the results.

MS. FLORES: Right.
CHAIRMAN DURHAM: Dr. Flores.
MS. FLORES: Yes.
You know, historically, competency-based education came up in the -- in the '70s because
the -- the usual bell curve test, and so it was thought that competency based -- well, we were really getting at these things that kids really needed, but I don't think that has happened. And I don't think we really are -- we're -- we're just kind of faking it with competency-based education. I think we need a lot of -- kids need a lot of knowledge, and I -- I didn't see that this test was very knowledge based. I saw very pedestrian language. It wasn't using language where kids would say oh, this is what this means, and -- and it was. I mean, I would think that probably an engineer or an accountant had worked on this -- on -- on the English language portion of it. It wasn't beautiful. It wasn't a beautiful test at all.

The math part was inscrutable -- absolutely inscrutable. And I took a lot -- a lot of math. So I just thought how is a little fifth grader, or a little seventh grader, you know, seriously going to do -- going to do that, and that's exactly what the engineers, you know, whom I met before when I was writing, they -- they -- they were really concerned about whether kids were really learning math, because they took homework to home, and these engineers could not help their children with the math. So I mean, I -- I think when we have people in the field who know and who say,
you know, this is -- this is not math. This is not what my kid, you know, should be knowing.

Two, I think that -- well, I mentioned the -- the historical part of getting away from high stakes tests, and that was a big -- a big portion of the discussions in education that we were having in the -- in the ' 70 s , and ' 80 s , and such. And then here we come to a system that is all based on this, and I -- I'm thinking about poor kids, minority kids, where we're not even thinking about strategies for them, for teachers, working with teachers.

Yesterday I was very disappointed, even with the administrators, and -- that we are -- are training to help the least -- the -- these most vulnerable kids, and I'm -- I'm just really worried. And especially, you know, given that we're not even thinking about what else we can do. And I'm sorry. Yes, it may show some truth, but the truth is very ugly, and we need to think of something other than what we're discussing here. And we need to really measure what those thing -- what those kids know. Give them a chance in some area maybe other than math that they can be successful in.

And I'm sure that the -- tomorrow's print will be that minorities did awful, and you know, the narrative continues. And I don't think this narrative is
a good narrative for those kids in need, and for really our society. These are not creative. They're not going to engender creativity in our society, and I think that's very sad.

And also, by the way, this -- these tests were really came about because of millionaires; Bill Gates to -- to be exact, who put in all this money to -- to start Common Core, and PARCC, so it has been driven -- I mean, it was a significant amount of money that was placed. And you can't say that because that it's not true, because it's true. So we have a case where billionaires are, in a sense, driving policy. We even got monies here to -- to help along in that -- in that respect. So I don't think it's a good thing.

Franklin -- Benjamin Franklin thought the common schools should be paid for by the common man, regular Joe, and not by rich people. And I think we're doing a great disservice to our public schools, and education in this state by allowing that to happen.

CHAIRMAN DURHAM: Thank you.
Any additional comments, Dr. Schroeder?
MS. SCHROEDER: If I may, I would just like to talk about one comment in your presentation. Each state will make its own accountability decisions. I personally think that's a pretty important piece of this,
not only to be -- as I said previously, to be cautious about this, but we have a number of options, and I -- so I'd just like to give an example that in another state, which I believe was Washington state, but I could be wrong.

I don't think they took a time out, but they changed the emphasis on the results. In other words, they said, we've only had these higher standards for $X$ number of years; therefore, as MS. Goff reminded us, a lot of the kids really never had been exposed to the standards, and so getting caught up is a long-term process. There isn't any way that we can take an 11th grader and have that 11th grader be up to par when, in fact, grades $\mathrm{K}-10$ they weren't exposed to those expectations.

And so they tried to find a metric that accommodated that, and honestly I can't remember whether it was grade-by-grade, or just exactly how it was done. I think my only point is that we do have a number of options, depending on whether we choose to take a time out, or not. We have different options so that this doesn't feel like an incredible hammer, based on the changes. And so I'd like us to think about that.

I'm not opposed to some -- some more time out, but I'm also -- even if we take one or two more
years, we nevertheless are going to have upper grade students, who have been trying to catch up, whereas we have early elementary grade kids, who from kindergarten on have been exposed to the standards, and have been expected to meet those standards, so that's a complexity of how you jump into -- this is across all the states -- how do you jump into higher standards, and have folks, and especially have teachers feel like this is really fair? So I just want to address that. CHAIRMAN DURHAM: MS. Goff, any additional comments?

MS. GOFF: No. Well, I would -- I would reemphasize the importance of our ability to do our pilots, and to develop our -- our approach to accountability, and do it in the best way we can. And then I believe we -- we will. And -- and without compromising the integrity of what Colorado knows is the right way to go. I do think all of our conversations around -- I hope you can hear me, I don't hear myself -- but all of our conversations around the demonstrations, and looking at our graduation guidelines, and the opportunities, and flexibilities within that framework. I -- I know I'm sounding probably way too umbrella-ish today, but that's how I tend to think. I just -- I just think we need to keep our eyes open always
for how to connect the dots, and how to supplement one -- one part of the system with another in the best way we know how.

I don't think this test, or any test -- I never have thought that this is the end all statement on how we're doing, and how kids are doing more importantly, but I do think we have what it takes to -- to -- to work through this, accept it for what it is. Recognize where we've got some stuff to do yet, and go at it in a reasonable approach that doesn't keep us lagging backwards with going down the road going forward the kids.

CHAIRMAN DURHAM: MS. Rankin, anything else?
MS. RANKIN: I just have a couple of questions. Your NAEP parallels with the ELA, the scores -- I'm sorry -- parallels with this test, as far as the scores, correct? How about math; did that do similar?

MS. ZURKOWSKI: Mr. Chair?
CHAIRMAN DURHAM: Yes.
MS. ZURKOWSKI: So when we look at the performance of our English learners on math, we do see that our students who are not English proficient, even with the accommodations, right -- so remember that those assessments can be translated into native language if
that's what those students are used to -- they are still performing lower than our native English speakers.

Did I misunderstand the question?
MS. RANKIN: I think maybe.
MS. ZURKOWSKI: I think so too, because that's what my note just told me is I missed that question.

So were you asking about mathematics?
MS. RANKIN: Yes. And -- but go back --
MS. ZURKOWSKI: I totally don't know what I heard.

MS. RANKIN: -- do you --
MS. ZURKOWSKI: So --
MS. RANKIN: I don't know what I said. ELA if -- if I understood you correctly, the ELA in -- in the NAEP, and in this test pretty much parallel --

MS. ZURKOWSKI: Uh-huh.
MS. RANKIN: -- with maybe a little bit of
adjustment. Which one was higher?
UNIDENTIFIED VOICE: This -- this test was a little bit higher --

MS. RANKIN: Little bit higher --
UNIDENTIFIED VOICE: -- by about five
points, was it?
MS. RANKIN: -- is that what you said?

MS. ZURKOWSKI: So --
MS. RANKIN: I'm not sure.
MS. ZURKOWSKI: -- in term -- Mr. Chair, apologize.

CHAIRMAN DURHAM: Proceed.
MS. ZURKOWSKI: So when we're looking at mathematics, we had a range between 19 and 37.

MS. RANKIN: Uh-huh.
MS. ZURKOWSKI: That 19 was at eighth grade, so I want to go back to talk about eighth grade again, because it -- that's really important what we have going on at eighth grade. For NAEP it was between 37 and 42. So our higher end was very close to what we had with NAEP. Our lower end, not so much. Again, remember that for our eighth grade results, that only includes our students who took the eighth grade assessment. It excludes our students who took algebra, or geometry, or integrated I, or integrated II students.

MS. RANKIN: That should be a lot actually. UNIDENTIFIED VOICE: Yeah.

MS. RANKIN: You should have algebra I in eighth grade.

MS. ZURKOWSKI: Mr. Chair?
CHAIRMAN DURHAM: Yes.
MS. ZURKOWSKI: So -- so as we're looking at
our eighth grade expectations, I would suggest that they're different than what we had when we were in eighth grade. It is -- is more reflective of some of those algebra I expectations that we may have experienced, but the students who are taking the eighth-grade test, like I said, are getting assessed on that eighth-grade content are eighth graders who are taking the algebra test are getting assessed on the algebra content. And it is fair to say that those populations of students aren't necessarily the same.

I -- we will encourage districts to look to see what past performance was of those students, and I think it is fair to suggest that our eighth graders who took the algebra test were our higher performers as seventh graders overall.

MS. RANKIN: So is that a yes or a no to my question?

MS. ZURKOWSKI: So it -- it was a well, kind of sort of. No. So as we're looking at our fourth graders, who on PARCC were at a 30 that is about, you know, seven to ten percentage points lower than what we had with NAEP.

MS. RANKIN: Okay.
MS. ZURKOWSKI: When we're looking at the eighth graders, I'm saying that --

MS. RANKIN: We can't --
MS. ZURKOWSKI: --it's -- it's messy.
MS. RANKIN: -- determine. I understand.

MS. ZURKOWSKI: It's --
MS. RANKIN: But even if you give me a third grade, it gives me an idea of the fact that they're parallel. And in ELA it was pretty much the same?

MS. ZURKOWSKI: Yeah.
MS. RANKIN: Okay. That helps a lot.
And then $I$ just want to go over, on page 15 -- you don't even have to look it up -- but it's level 1 through 5 --

MS. ZURKOWSKI: Yep.
MS. RANKIN: -- you remember the levels you gave? When I'm talking to a parent -

MS. ZURKOWSKI: Yes.
MS. RANKIN: -- is it going to be appropriate to say yeah, a number 1 is an $F$, number 2 is a D? Tell me why not.

MS. ZURKOWSKI: Mr. Chair?
CHAIRMAN DURHAM: Yes, please proceed. I want to hear this one.

MS. ZURKOWSKI: I'm -- I'm glad that
my -- my -- my head shake was registered. So as we -- again, I think that depending on the school that
you're in, interpretation of an $A, B, C, D, F$ is very, very different. There are some schools -- again, looking back at some of our own experiences that we do, I think, sometimes as we're sitting in these chairs, some folks used to grade on the curve, and so you would always just have 16 percent of your kids who were an A. And then at a B level you would have an additional 30 percent of your kids who may have scored there. And then at a C - that is not how this is done.

What is fair to say is that at a level 4 and a level 5 students are demonstrating those concepts and skills that we would expect them to be able to demonstrate in order to be successful at their next spot, whether that be the next -- next grade level, or the next -- for college, or careers, but it's not an A, B, C, D, F kind of concept.

MS. RANKIN: Okay.
CHAIRMAN DURHAM: Further questions, MS.
Mazanec?
MS. MAZANEC: Couple of more questions. What is -- on one of these slides you have -- you have FELL/PHLOTE.

MS. ZURKOWSKI: Yeah.
MS. MAZANEC: Would you explain what those mean?

MS. ZURKOWSKI: Absolutely. Sorry. So we --

MS. MAZANEC: The fluent --

MS. ZURKOWSKI: -- have -- our students --
MS. MAZANEC: -- fluent English as the language learner.

MS. ZURKOWSKI: -- so our students who come into our schools one of the questions that is asked is, what are the -- what's the language that you speak at home, you know, what's the -- that home language. For students how have a home language other than English, they have a primary home language other than English, but they don't require English learner services, they're referred to as PHLOTE, primary home language other than English. Again, complicated.

MS. MAZANEC: And FELL is --
MS. ZURKOWSKI: If a student -- yep, so if --

MS. MAZANEC: -- fluent English language
learner?

MS. ZURKOWSKI: So for FELL those are our
former --

MS. MAZANEC: Oh, former.
MS. ZURKOWSKI: -- English language
learners. So those are students who, at one point in
time, were English learners. They are no longer designated as English learners, i.e., they've become proficient in English.

MS. MAZANEC: One -- one other question. I know that we're not -- we're not to compare these test results to TCAP results, but remind me with our last round of TCAP did we not have better performance among elementary students then, as -- as opposed to high school students too?

MS. ZURKOWSKI: So -- and you'll notice that I glanced over at Dr. Asp here for a second.

MS. MAZANEC: Dr. Elliott.
MS. ZURKOWSKI: But you definitely saw a difference in performance, especially when we looked at mathematics, in terms of what we had in terms of performance of our elementary students compared to our high school students. And yes, our elementary students were performing higher than our high school students. I don't know if you want to add to that. MS. MAZANEC: Has that been true for some time?

MR. ASP: And I -- I try not to get real technical about this, but it's a great question. In -- in reading and writing the differences between elementary, middle, and high school were not big at all.

There's percentage of kids proficient.
In math, the way in which the -- the
performance levels -- the -- the what was proficient was set at a much different way, so they were deliberately made much harder, even though it's harder to be proficient in math in high school than it is in -- anyway, it -- it was pumped up even higher, so those -- it always resulted in that similar pattern: high elementary, little lower middle, little lower high school, because the performance levels were harder.

MS. MAZANEC: Thank you.
CHAIRMAN DURHAM: All right. Let me --
MS. SCHROEDER: May I ask --
CHAIRMAN DURHAM: -- we'll go back around if we need to. Dr. Schroeder -- Dr. Scheffel, I'm sorry.

MS. SCHROEDER: Yes.
MS. SCHEFFEL: Yeah, as I think
about -- this is a great conversation. Appreciate all the great information. It seems to me that we need to ask the question for whom are we doing these tests. Is it for the students, for the adults? What is it for? And right now we're doing it for high stakes sorting of districts for accreditation, and also wondering about how else we can use these data for teach evaluation, or whatever. And we're using a test that's not content based. Not based on what students know, but on language really -- language acumen. And it uses a very subjective approach to the proficiency levels, and how those bands are set in order to sort. And so I -- I think we know, as educators, that a high quality education is -- is characterized by students doing things like science experiments, solving real world math problems, writing research papers, reading novels, and stories, and analyzing them, making oral presentations, evaluating and synthesizing information from a variety of fields, and applying that information. And standardized tests are just not a good tool for evaluating that kind of learning.

You know, if we look at Finland, many nations that do the best in international comparisons do not use wide-scale standardized assessments. So if we know that instruction is really trying to teach to the test, and we have identified the biggest factor loadings on the test, then I -- I don't - I don't know that students are going to be able to learn the kinds of skills, and engage in the kinds of activities that we've just listed that we know comprise of quality education.

So I think we have some soul searching to do. If our biggest goal is to be able to sort districts
for accreditation, I -- I think that we're leaving the kids behind, because it's not necessarily meeting their needs. So I -- I think this is a great discussion, and I know it's the first cut on the data, but I think we really have to return to our basic questions, you know, for whom are we giving these tests, and how are we intending to use these data.

Thank you.
CHAIRMAN DURHAM: Dr. Flores.
MS. FLORES: And -- and we really should ask what is the purpose of education. You know, I think that's -- that's what we should be asking, what is the purpose of education. And I don't think -- I don't think it's about accountability. I don't think it's about testing. I think it -- it really gets down to education is about teaching and learning. And we are so far removed from that.

I think too -- I mean, I thought -- I
thought this for a long time, even before we had those other tests, that if Cesar wants to know -- if Cesar wants to know how kids are doing out in the country, just give it to him. And what does he have? He has -- or she has NAEP, National Assessment for Educational Progress, which you said was almost the same, and that's what is used nationally. Give that to him. Let's not get
ourselves into spending more money in -- more money in -- in this accountability stuff. It -- they -- they want it. They get it. They have it through NAEP. They do it every year.

So I think we need to start thinking about that, and then think about local. Local assessments are better than these high stakes testing. They will do more. And I just think when you have 7 states that are cooperating with PARCC, you have 43 states that said no -- no to PARCC. And we should really think about that. Why should we be the experiment? I don't like to be -- I don't want to be. I don't want our kids to be the experiment.

Thank you.
CHAIRMAN DURHAM: Joyce, I'll try and conclude with a couple of things. One, you and I have had a number of discussions about -- about the potential arbitrary nature of using -- of -- of the -- of the criteria based standards being set, based on a group of individual's judgment of what people should know. And that is a completely arbitrary -- it may be well intentioned, but you cannot take away from the fact that it's the arbitrary judgment of a group of people.

So one of the things that $I$-- we discussed, and I asked is whether it's possible to convert these
scores on a district-wide basis and a school-wide basis to a norm referenced -- a norm-referenced basis. And I think we concluded that that could be done. That you know what every individual scored. You know what every individual scored in the school. You can get an average for each school. You can get an average for each grade in each school, and each district.

And I don't -- with modern technology, I don't believe that's too hard to do, and so I'm going to ask for a consensus from the Board to -- to ask you to do that, and provide those to us, and to the public at the next meeting. So that the public can still compare school $A$ with school $B$ and district $A$ with district $B$, and allow them to make choices. There are a number of private and public organizations that rely on this kind of data.

The -- the criticism of -- of the norm-based data is that -- I had to find some way to politically correctly say this -- is that we're trying to find that all we're really finding is the fastest turtle, because we don't compare ourselves with Singapore, and - and Finland that it's really not relevant data for worldwide comparison. And that may or may not be true, but it's certainly -- it's certainly relevant, so that we can compare ourselves internally. And I'm going to guess
that there are -- there's probably data available from the other states that they cannot hide. They're probably publishing it this morning, as well, that would allow us to compare ourselves with the six other PARCC states. And -- and I think when you -- when you evaluate the effectiveness of -- of these exams I think you have to look at a couple of things. Only six states use the PARCC test. Number often heard is seven, but the last time $I$ checked -- and of course I have old information, the District of Columbia is not a state. Fifteen states use Smarter Balance. So by my old math that's 21 states that have acquiesced in the beginning, or continue to acquiesce to the attempt at federalization of these testing consortiums. At least 25 states, and that leaves a few states in doubt, use state-specific assessments in grades three through eight in math in English arts.

So quite clearly, we are in a very small minority, which -- and -- and I think the whole concept -- the reason that we went down this road, starting with no child left behind, was that somehow we needed to be able to compare our performance -- the performance of our students with Massachusetts and Mississippi. I don't think that's even remotely possible with the -- the fact that we now have, at least, 27
different -- different standardized tests out there trying to draw these conclusions, and whether they can be successfully normed one to another, I have my -- my serious doubts.

I think -- I do think the other element that -- that's been alluded to that these tests don't test knowledge, and they certainly don't try and test any common body of knowledge that we would expect students, as Americans, to know or understand, and that I think is a serious shortcoming. When, in the sample junior high question you have here, no one is asked to draw on any knowledge of literature they have read, or should have read to be able to -- to be able to answer the question. So they're -- they're by and large knowledge absent tests. And I think that is -- is -- is part -- is one of the fundamental problems.

The other fundamental problem is -- is next year's cut scores could be subject to manipulation, just as this year's cut scores could have been subject to manipulation, because you'll have new questions. You will have new evaluations of how many of those questions someone could get right. You will probably have new evaluators. I don't expect you will have the same 30 or 40 experts setting the standard of knowledge that we expect sixth graders to know when they take the math
test. So I'm not, at all, convinced that by using -- by using this criteria referenced approach, that we will ever have tests that can be compared from year-to-year, and really be an accurate portrayal of academic progress.

So I think -- I think there are a couple of things that we can hope for. If -- if you take these results at face value, almost two-thirds of our students have failed.

UNIDENTIFIED VOICE: They have.
CHAIRMAN DURHAM: Now, I have a hard time believing that number, because we have two-thirds of our students who have failed to attain what we believe they need to know to move to the next grade, which means we ought to be retaining 60 percent of our students in the grade they're in. If those are the results, you can -- you can draw very few -- your number of conclusions you can draw are very limited.

One of them is, our school system is a catastrophic failure. It may be. It may be that we're the slowest turtle. And on that measurement, we are a catastrophic failure. I'm not inclined to believe that these results are as dire as they appear, but there is no way for me to assert that with any greater degree of certainty than anyone else on this -- on this dais can assert that they are, in fact, an accurate and true
measure of what's going on, and that's the fundamental problem with these tests.

So the -- the other problem I think -- I think Dr. Scheffel clearly mentioned is, who are we trying to serve. Are we trying to serve children? Because if we're trying to serve children, these results would be available in a timely fashion, and we should be able to help them improve, because they have the results available so the teachers know what they need to know. We're almost halfway through, and by the time these results trickle down it'll be Christmas vacation. So for all intense and purposes, we're halfway through another school year without the -- the helpful commentary, that can be associated with these exams, available to the teachers that have inherited these students after -- after they took the tests last -- last May. So from that perspective, these tests have failed to serve the one group that we should be trying the hardest to serve, which is the children of the State of Colorado. And in that regard, they're a catastrophic failure, and -- and if we proceed with PARCC, we need to insist that these results are available in a much more timely fashion. That -- the absolute minimum we ought to be able to achieve.

So I would -- would anticipate, if I don't
see objection, that we will have the norm-based results comparable for districts, and I would presume that we should be able to -- if someone scored 650 on the test, we should probably be able to convert that to, at least, a state-wide average of what percentile that is.

So with that, I think we'll conclude.
Oh, yes, Dr. Asp.
MR. ASP: (Indiscernible).
CHAIRMAN DURHAM: No, go ahead.
MR. ASP: We can certainly produce those percentile results. Having those for you in December will be very difficult for us to do, given the other pieces that we need to produce for that.

MS. FLORES: But don't you have those, so they'd convert them to this? And one last -- last thing. Why are we even thinking of pursuing any further -- any further use of this test? I think we should be thinking about getting out of -- of -- getting out of PARCC.

CHAIRMAN DURHAM: I think the answer -- the answer, Dr. Flores, is that the law requires us to use it, and whether we like it or not, we're going to -- going to have to follow that law. And until -- until the policy makers conclude that, you know, either this is the right way, or this is a flawed process --

MS. FLORES: It's a flawed process.
CHAIRMAN DURHAM: -- we're stuck. So --
MR. ASP: We'll -- we'll look this, and give you --

CHAIRMAN DURHAM: -- and --
MR. ASP: -- and get back (indiscernible).
CHAIRMAN DURHAM: -- and Dr. Asp, I
would -- I mean, I -- you know, my computer knowledge is not the greatest, but it -- it would seem, to me, that -- that it is maybe -- maybe I oversimplified it as a one button push on a computer, but maybe it's two buttons. I don't know.

MR. ASP: We'll do the best
(indiscernible) --
MS. ZURKOWSKI: Give me five.
CHAIRMAN DURHAM: But -- you know, and -- and I -- I would simply say we have other obligations that we need to meet before we provide the -- the normed-based data, so those will come first, and -- and with the limitations of staff and time, but as soon as possible I think it's going to be very important for really everyone in this state to be able to start to make those cross district and school-by-school comparisons.

So thank you very much.

MS. FLORES: Get out of PARCC. I would like to have our public have the assurance that this will be clear. There's -- there's quite a bit of difference in understanding what it means to report in this -- in this way we have versus the other way. And if you've got -- if you -- if we're going to do this, I would just adamantly ask that there be some way of outlining for the public and parents what the difference is here, and what it means, what this number means, as compared to this number, if you can do that. I -- I (indiscernible) --

CHAIRMAN DURHAM: Your catch phrase is when finding the fastest turtle you use this particular data, so all right.

Why don't we stand -- why don't we go ahead and proceed to our lunch break and executive session. We will lay over the rule making hearing until the first thing after lunch at about quarter till 1:00. And we'll stand at recess until 12:45 p.m.
(Meeting adjourned)

C E R T I F I C A TE
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 25th 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

