

UNITED STATES DEPARTMENT OF EDUCATION

LESSONS LEARNED FOR IMPROVING ACCESS TO GENERAL
ASSESSMENTS BY LOW-PERFORMING STUDENTS WITH
DISABILITIES

Washington, D.C.

Tuesday, May 22, 2012

1 PARTICIPANTS:

2 U.S. Department of Education:

3 MELODY MUSGROVE
4 Director, Office of Special Education Programs

5 ALEXA POSNY
6 Assistant Secretary, Office of Special
7 Education and Rehabilitative Services

8 PATRICK ROONEY
9 Race to the Top Assessment Team Lead
10 Implementation and Support Unit

11 SUSAN WEIGERT
12 Office of Special Education Programs

13 ANN WHALEN
14 Director of Policy and Program Implementation
15 Implementation and Support Unit

16 Experts:

17 LOU DANIELSON
18 American Institutes for Research (AIR)

19 STEPHEN ELLIOTT
20 Founding Director, Learning Sciences Institute
21 Arizona State University
22 Mickelson Foundation Professor of Education,
Mary Lou Fulton Teachers College

STEVE FERRARA
Vice President of the Center for Performance
Assessment Pearson Assessments

SHERYL LAZARUS
Senior Research Associate, National Center for
Educational Outcomes
Co-Principal Investigator, University of
Minnesota (NCEO)

1 PARTICIPANTS (CONT'D):

2 SHELLY LOVING-RYDER
Assistant Superintendent for Student
3 Assessment and School Improvement Virginia
Department of Education (VDOE)

4 Smarter Balanced Assessment Consortium:

5 WENDY CARVER
6 Utah State Office of Education

7 GAYE FEDORCHAK
New Hampshire Department of Education

8 MARGARET HO
9 Washington Office of Superintendent of Public
Instruction

10 MICHAEL HOCK
11 Vermont Department of Education

12 DEBORAH MATTHEWS
Kansas State Department of Education

13 JENNIFER PAUL
14 Michigan Department of Education

15 JOE WILLHOFT
Smarter Balanced Assessment Consortium

16 Partnership for Assessment of Readiness for
17 College and Career (PARCC):

18 ROBERTA ALLEY
Arizona Department of Education

19 TRINELL BOWMAN
20 Maryland State Department of Education

21 ANDREW HINKLE
Ohio Department of Education

22

1 PARTICIPANTS (CONT'D):

2 JEFF NELLHAUS
3 Achieve, Inc.

4 DAN WIENER
5 Massachusetts Department of Elementary and
6 Secondary Education

7 LEILA WILLIAMS
8 Arizona Department of Education

9 National Center and State Collaborative (NCSC):

10 RACHEL QUENEMOEN
11 National Center on Educational Outcomes (NCEO)

12 Dynamic Learning Maps (DLM):

13 ALAN SCHEINKER
14 Center for Educational Testing and Evaluation
15 University of Kansas

16

17 * * * * *

18

19

20

21

22

1 P R O C E E D I N G S

2 (8:30 a.m.)

3 MS. POSNY: Everyone's been so open
4 about talking and sharing our ideas and things
5 that we're working together, and I think it makes
6 an incredible difference when we think about it,
7 because this is really and truly a partnership,
8 because we really and truly want to include all
9 students when we think about the assessments and
10 where we're headed in our next generation. So,
11 this is truly a partnership between and among
12 OSERS, the Implementation and Support Unit in
13 which Ann is in charge, as well as the key state
14 school officers in terms of ASES and SCASS.

15 And you know, whenever I talk about
16 these and I put in all of these alphabets, people
17 look at me and they say so what does it all mean?
18 And I say all it means is that we are assessing
19 all kids, so think about it that way.

20 And of course we want to thank the
21 people who helped fund this initiative for us as
22 well.

1 Okay, the goals. The goals are really
2 pretty straightforward in terms of what we really
3 want to do. You know, we in OSHA's Special
4 Education Office had funding for a number of years
5 -- the last 5 or 10 years, and, Larry, you should
6 be the one talking about this, but we have had
7 funding for a number of years, what's referred to
8 as the GSEG. And the GSEG is really -- it's kind
9 of the beginning of how we began to assess
10 students with disabilities. We've learned a lot
11 over the 5 to 10 years. I don't know about you,
12 but I've learned a heck of a lot. And it's really
13 -- we've really come a long way.

14 The whole idea behind the meeting today
15 is what have we learned and what can we use in
16 terms of developing this next set of assessments?
17 And what we want to do is we really want to offer
18 assistance to any and all in terms of going
19 through this so we don't have to really start from
20 scratch all over again. We've people who are
21 extremely knowledgeable about how we can make all
22 assessments accessible for the majority of our

1 kids.

2 What we're going to do is we're going to
3 be hearing from five of our former grantees.
4 They're going to present their research and the
5 lessons that they've learned over the course of
6 that time and how they did work on the
7 accessibility of all assessments, especially for
8 students with disabilities. And know that a lot
9 of what we're talking about today will also apply
10 toward English-language learners as well. We also
11 then want to hear from the discussants who
12 represent all the -- the other four consortia from
13 PARCC, Smarter Balanced, the National Center and
14 State Collaborative, and the Dynamic Learning
15 Maps. We're going to open it up and have them ask
16 questions, because this is a chance for all of us
17 to learn from each other.

18 When we think about how long this has
19 been going on, think about the consortia in terms
20 of PARCC and Smarter Balanced. These were awarded
21 in September 2010, and then we in OSHA awarded
22 ours to the two consortia shortly thereafter,

1 these four-year grants to develop the next
2 state-of-the-art in terms of the statewide
3 assessments for the first time with the consortia
4 across a majority of the states.

5 When we think of the GSEG, the General
6 Supervision Enhancement Grants, funded by special
7 grant, these are really for the assessments for
8 those kids with the most significant cognitive
9 disabilities. This is what we have always
10 referred to as the 1 percent, because the Race to
11 the Top Assessment is going to assess 99 percent
12 of all other kids.

13 When we go back to the GSEG, we know
14 that we have two consortia who are one of those.
15 It's comprised of 31 states as well as Washington,
16 D.C., and again this is the National Center and
17 State Collaborative and the Dynamic Learning Maps.

18 The Race to the Top, two consortia
19 representing 25 states and D.C., and this is PARCC
20 and Smarter Balanced. So, when we give out the
21 picture of assessments in 2014 and 2015, it's
22 going to be these four assessment consortia that

1 will really assess the majority of kids across
2 this country.

3 When I think about IDEA, Individuals
4 with Disabilities Education Act, I look at this as
5 a true partnership between (inaudible) general and
6 special education.

7 Whenever I think about special ed and I
8 think about students with disabilities -- students
9 with disabilities, they're general education
10 students first and foremost. When you think that
11 more than 60 percent of our students with
12 disabilities are in the general ed classroom more
13 than 80 percent of the day, who are the primary
14 educators of our students with disabilities? It
15 happens to be the general educators.

16 Critically important is when you think
17 about the future of kids with disabilities. For
18 too long, we expected too little from kids with
19 disabilities, and what they have done, I think, is
20 nothing short of phenomenal. For the last the
21 increases we have seen in both academic
22 performance and what they're capable of doing has

1 been not short of wonderful.

2 When we think about kids who are served
3 under IDEA, and I'm talking about kids from the
4 age of birth to the age of 21, we serve all 7
5 million kids with disabilities in the public
6 schools. Compared that to 1974 when almost 2
7 million kids with disabilities were completely
8 excluded from school and never entered the school
9 door. I'd say we've done a great job in terms of
10 making sure that we educate all kids.

11 When we think about the 0 to age 3, this
12 is absolutely 3 percent of the population that we
13 serve, and when we think about students with
14 disabilities in the public schools, we're talking
15 about 14 percent of the population. We categorize
16 them into 13 different disability categories. The
17 majority of them, 41 percent, are classified as
18 having a specific learning disability. What I
19 want you to understand in these categories, if
20 you're not familiar to it, even though I know I'll
21 probably try to require more than anything else,
22 is we have a full range under a child who is LD.

1 Could be very mild, could be very moderate, could
2 be very severe. So, within each category we have
3 a range of severity in terms of the disability.

4 The next group is speech/language
5 impairment, and this is not just a speech
6 impediment. It includes language skills, which
7 are critically important when we consider access
8 to the printed word.

9 Seventeen percent of all other
10 categories, including other health impairments or
11 the blind and visually impaired and deaf and hard
12 of hearing are in the 17 percent; 8 percent are
13 classified as having intellectual disabilities; 7
14 percent emotionally disturbed; and 6 percent with
15 autism. Two percent of the population have what
16 we refer to as multiple disabilities, meaning they
17 have more than one. So, you can kind of get a
18 flavor for when we think about assessing 99
19 percent of all kids within the assessments. We're
20 talking about a very, very diverse group of kids
21 that are included.

22 So, over the course of time, honors

1 students with disabilities participated in
2 statewide assessments. For some, this is a rather
3 surprising graph, and yet, to me, it makes perfect
4 sense in terms of the timing and in terms of we've
5 learned over the course of time.

6 In 2007 and '08, we had a little over
7 percent of kids with disabilities who participated
8 in what we refer to as alternate assessment. I
9 want you to understand that in that alternate
10 assessment, we're not just talking about the 1
11 percent; we're talking about the 1 percent and the
12 2 percent at that point in time. So, what you've
13 seen is an increase in terms of the number of
14 students with disabilities that are in the
15 alternate assessment. It's because of the
16 approval of more of the 2 percent assessments, and
17 that's what we're seeing. But what we're going to
18 see over the course of time is that we're going to
19 see this change, because we are talking about only
20 1 percent of the population being assessed with
21 alternate assessment. So, over the course of
22 time, we expect and anticipate to see probably at

1 least 90 percent of all kids with disabilities be
2 assessed with the general assessment, and that's
3 the key as to why we're here today.

4 The thing that has driven us probably
5 more than anything else, and what we have seen
6 over the course of the last 10 to 12 years is that
7 students with disabilities can meet college and
8 career learning standards. We didn't know that
9 until we started assessing them. And what I want
10 you to remember is that the assessment of kids
11 with disabilities did not occur with the last
12 three authorizations in 2002. The assessment of
13 students with disabilities occurred in the 1997
14 reauthorization. It was at that point in time we
15 were required to assess all kids with
16 disabilities. At that point, we just didn't know
17 how, and that is what we have been working on over
18 the last 15 years.

19 We also know that students with
20 disabilities can and do excel within the general
21 curriculum, and we must prepare them for success
22 in postsecondary education as well as in career.

1 We have to include kids with disabilities if we
2 are going to meet President Obama and Arne's goal
3 about having the highest number of college
4 graduates across the world. So, in order to do
5 that, we must ensure that all kids with
6 disabilities have access to the general curriculum
7 and that we do everything that we can to intervene
8 as much as possible to meet their unique needs to
9 make sure that they are successful. They also
10 know that IEPs must be a biodegradable academic
11 standards so that we are making sure that kids are
12 learning to the same standards as all others. And
13 we know that we need the best and most effective
14 teachers that we can possibly have. And again
15 remember that other statistic that I shared with
16 you. We must have general educators who know how
17 to work with and teach effectively all students
18 with disabilities, along with special educators,
19 so that everyone can individualize the
20 instruction.

21 The last thing I want to share with you
22 is what kind of progress has been made, and this

1 is based upon the innate reading and math scores,
2 as well as our most current data that we have. In
3 terms of all students, an increase in the percent
4 who are at proficient and above, for all students
5 it was 8 percent. For students with disabilities,
6 it was a growth of 8 percent. In terms of math,
7 slightly higher for the general population with a
8 6 percent increase for students with disabilities;
9 that would be 8th grade.

10 Look at the graduation rate. For the
11 general population, it's increased by 3.4 percent,
12 for students with disabilities by 10 percent.
13 That's huge. And when we think about it, the
14 statistic that I keep in mind is almost 60 percent
15 of students with disabilities graduate with a
16 diploma. That's in comparison to 66 percent of
17 the general population. We're still behind, but
18 at least we're catching up or headed in the right
19 direction.

20 The one that I find the most profound is
21 look at the increase in postsecondary enrolling.
22 If you were to talk the facts to juniors and

1 seniors, kids who have disabilities in the high
2 schools, and you had asked them what they want to
3 do and where they're going to go on next, almost
4 80 percent or more of them are going to say I want
5 to go on to college. They all have that
6 expectation. Ten years ago that would not have
7 been the case. So, we now have 38 percent of kids
8 with disabilities who go on to some postsecondary
9 enrollment. And 4-year college enrollment -- just
10 15 years ago, a 4-year college enrollment in kids
11 with disabilities was less than 1 percent. Now,
12 we're up to 13 percent. We still have a ways to
13 go. But at least, again, we're headed in the
14 right direction.

15 With that, I'd like to turn it over to
16 Ann.

17 MS. WHALEN: Thank you. Good morning,
18 everybody. Just wanted to thank everybody for
19 traveling from across the country to come to
20 today's meeting and, more importantly, to get
21 through our PCP security. I apologize that it's
22 such a hassle, and I know many of you waited in

1 the lines that circled around the building, so we
2 appreciate that.

3 We are lucky enough with the support of
4 the Hewlett Foundation to help support meetings
5 like this today, and this may seem minor, but
6 they're able to help get us water and coffee,
7 which if you've ever been to a federal meeting
8 before is a luxury. So, please enjoy it while we
9 can.

10 But more importantly, they made it
11 possible for us to come together as a community
12 and get smarter and learn from our grants that
13 we've had an opportunity to be in place for a
14 couple of years but also learn from each other.
15 So, this is really a chance to support the
16 collaboration of base multiple consortia, states
17 coming together to develop and design and
18 implement the next generation of assessments.

19 So, just as a quick reminder -- not on
20 this slide, sorry. So, just as a quick reminder,
21 the Race to the Top Assessment grant was a
22 state-led effort, and needs to be a state-led

1 effort, to develop and deliver the next generation
2 of assessments. And these must provide accurate
3 information to our students no one can do, reflect
4 good instruction practices, and support a culture
5 which needs improvement, and effectively it
6 affects all students, including students with
7 disabilities and English-language learners.

8 When we're looking forward, these
9 assessments must include one or more summative
10 components and be able to be delivered by the
11 2014/15 school year, and for some of us that feels
12 like into the future. For many of the people at
13 this table and in the audience, that's really,
14 really close. But we're really excited about this
15 work. But at a minimum these assessment systems
16 must be administered at least once during the
17 academic year and at a minimum include
18 English-language, arts, and mathematics in reading
19 grades 3 through 8 and at least once in high
20 school. And, more importantly or just as
21 importantly, these results are, from the
22 beginning, designed to be used to inform teaching

1 and learning and program improvement;
2 determination of school effectiveness;
3 determination of principal and teacher
4 effectiveness for the purposes of evaluation and
5 support; and determination of individual students'
6 college and career readiness.

7 For today's conversation, we really
8 wanted to highlight some of the ways from the
9 beginning that we're trying to include students
10 with disabilities in the Race to the Top
11 Assessment, design, development, and
12 implementation. So, part of the absolute priority
13 -- we require that the states create assessments
14 for all students, including students with
15 disabilities and English learners, that the
16 consortia are required to develop tests accessible
17 for these populations and create standardized
18 accommodation policies, and both PARCC and Smarter
19 Balanced have established panels of external
20 experts knowledgeable about the needs of students
21 with disabilities to inform assessments and
22 development.

1 Again, I just want to thank everybody so
2 much for your time and energy today and your
3 willingness to not just come and share what's
4 working but, just as importantly, come and share
5 what's not working, because it's very important
6 that as a community we continuously improve and
7 get better upon ourselves.

8 So, with that I'm going to turn it over
9 to Susan to walk through a little bit of the
10 logistics for the day.

11 MS. WEIGERT: Thank you. So, I just
12 wanted to go over the agenda. We are going to
13 start our speakers at 9 o'clock or just where they
14 get cut, and then the public part of the meeting
15 will be over at noon, and we're going to have a
16 private meeting with the consortia for consortia
17 business at 1 o'clock. And so they will stay on
18 till 2 o'clock. And then many of you will be
19 going over to the ASES meeting at about 2 o'clock.

20 So, I just wanted to introduce our
21 invited experts. There's Lou Danielson from the
22 American Institutes for Research, and Lou is the

1 former OSEP director, as many of you know; Steve
2 Elliott is here from Arizona State University.
3 Steve Ferrara -- Steve, why don't you come on up
4 and join us here -- Steve's from Pearson; and
5 Sheryl Lazarus from NCEO; Shelley Loving-Ryder
6 from the Virginia Department of Education. These
7 are people who all have worked on the alternate
8 assessment on modified achievement standards under
9 the former 2 percent GSEG, and they're eager to
10 share some of their lessons learned about
11 development and assessment items that are
12 accessible to this population.

13 Just as a reminder, if you would place
14 all your cell phones and other devices on vibrate.
15 We have several resources in PowerPoint. And
16 after each of the speakers speaks, they will open
17 it up to questions from the table. At the end of
18 the meeting I will show you how you can write in
19 questions also that can be asked and answered.

20 So, with that, I will turn over the
21 pointer here to our first speaker.

22 MS. WHALEN: So, while we're doing that,

1 can we just go around the table so everybody has a
2 sense of who's here --

3 MS. WEIGERT: Sure.

4 MS. WHALEN: -- and what group or
5 consortium do they represent?

6 MS. WEIGERT: Why don't you go -- okay,
7 go ahead.

8 MS. WILLIAMS: Hi. I'm Leila Williams
9 with Arizona, and I'm here also on behalf of
10 PARCC.

11 MS. HO: Margaret Ho, State of
12 Washington, Smarter Balanced.

13 MS. ALLEY: I'm Roberta Alley of Arizona
14 and from Leadership (inaudible).

15 MR. HOCK: I'm Michael Hock from
16 Vermont.

17 MS. WHALEN: Excuse me, we can't hear
18 back here at all. Thank you.

19 MR. HOCK: I'm Michael Hock from Vermont
20 and I'm working with the Smarter Balanced
21 (inaudible).

22 MS. WHALEN: Can't hear.

1 MR. HOCK: I'm Michael Hock from
2 Vermont, and I am working with the Smarter
3 Balanced Assessment Consortium.

4 MS. BOWMAN: Good morning everyone. I'm
5 Trinell Bowman and I'm with the PARCC Consortium.

6 MR. DANIELSON: Lou Danielson from AIR.

7 MS. LAZARUS: Sheryl Lazarus, NCEO.

8 MR. FERRARA: Steve Ferrara from
9 Pearson.

10 MS. QUENEMOEN: Rachel Quenemoen from
11 NCSC and NCEO.

12 MR. SHEINKER: Alan Sheinker, Dynamic
13 Learning Maps.

14 MS. WEIGERT: Susan Weigert from OSEP.

15 MS. WHALEN: Ann Whalen from the
16 Implementation and Support Unit.

17 And just so everybody knows the rules,
18 push to talk. If your red light's on, that means
19 you can talk. We can only have one on at a time.
20 If you have cell phones, please remove them from
21 the table, because they will set things ablaze.

22 Ann, please let people on the phone know

1 that they're working on the slides (inaudible).

2 MS. WEIGERT: Yes, I'm so sorry, there
3 are people on the Webinar. We are working on the
4 technology difficulties with the PowerPoint
5 slides. We're hopefully going to get that settled
6 in a couple of moments. We apologize. But just
7 so everybody knows, all of this information will
8 also be available on our website starting today.
9 So, we will do our best to get everything up as
10 soon as possible but get it cleaned it up as we go
11 along. Thanks.

12 MR. ROONEY: Patrick Rooney, and I'm
13 with the Department of Education. I also work on
14 the Race to the Top Assessment team.

15 MR. WILLHOFT: I'm Joe Willhoft, and I'm
16 with the Smarter Balanced Assessment.

17 MS. MATTHEWS: I'm Debbie Matthews. I'm
18 from Kansas, and I'm here for Smarter Balanced and
19 DOM.

20 MS. LOVING-RYDER: Shelley Loving-Ryder,
21 Virginia Department of Education.

22 MR. ELLIOTT: Steve Elliott, Learning

1 Sciences Institute, Arizona State University.

2 MR. WIENER: I'm Dan Wiener from
3 Massachusetts. I'm representing PARCC. We're
4 also members of the NCSC Alternate Assessment
5 Consortium.

6 MR. HINKLE: Andrew Hinkle, Ohio PARCC.

7 MS. CARVER: I'm Wendy Carver from Utah.
8 I'm representing Smarter Balanced and the DLM
9 alternate assessment.

10 MS. FEDORCHAK: I'm Gaye Fedorchak. I'm
11 from the state of New Hampshire, and I'm
12 representing Smarter Balanced Assessment
13 Consortium.

14 MS. PAUL: I'm Jen Paul. I'm from
15 Michigan, and I'm here for Smarter Balanced.

16 MS. POSNY: And I just want to introduce
17 Melody Musgrove. Melody, you want to give a wave?
18 She is the director of the Office of Special
19 Education Programs.

20 MS. WEIGERT: All right. With that, we
21 will start our first speaker, Lou Danielson.

22 MR. DANIELSON: Thank you, Susan, and

1 thanks, Alexa and Susan for inviting me.

2 I'm going to begin providing a little
3 historical perspective, because I think sometimes
4 in our discussions we tend to focus on the
5 challenges and issues, and we forget how far we've
6 come, and I think it's important to acknowledge
7 how far we've come in this work.

8 The work related to -- including the
9 displaced assessments, broke down in the early
10 1990s, and one of the very first things that OSEP
11 did was to fund the National Center on Educational
12 Outcomes, which has now just completed about 20
13 years' operation, has begun I guess must be his
14 fifth funding cycle. And of course that's been a
15 critical investment, working to assist in
16 advancing our understandings in this area in
17 conducting some research and also, more
18 importantly, providing a lot of take-home
19 assistance across the country.

20 At the same time as the initial work was
21 funded, NCEO also funded -- began funding major
22 research investments, and I think now with 20

1 years- plus work in accommodations research,
2 there's now a good deal of research that exists
3 related to providing accessibility to kids with
4 disabilities, as well as developing principles of
5 universal design, which of course considerations
6 hopefully prevent the need for accommodations for
7 kids with disabilities.

8 In '96, an important breakthrough was
9 when NAEP began providing accommodations in the
10 National Assessment of Educational Progress and
11 began then, also at the same time, a new trend
12 line so that the data that Alexa presented in
13 terms of trends over time and improvements in
14 performance gives (inaudible) disabilities as
15 possible now to report that data.

16 And importantly, then, in 1997 was the
17 inclusion of provisions in the authorized IDEA to
18 require the state assessments to include students
19 with disabilities and also was when the
20 requirement for an alternate assessment first came
21 about in the main value -- and one of the very
22 important aspects of that is that between the two

1 assessments there was the alternate as well as the
2 regular assessment and expectation that all kids
3 with disabilities would be included in state and
4 local assessments. Sadly, there is not an
5 alternate with the NAEP, and so one of the things
6 we see with NAEP still is the fairly substantial
7 population of kids with disabilities who are
8 excluded in the NAEP.

9 In 2001, then, NCLB was enacted. And,
10 in my view, one of the things that's very
11 important about that is that for the first time we
12 see a piece of general education legislation that
13 kind of also codifies this expectation of kids
14 with disabilities being put in assessment. And
15 NCLB really kind of extended the IDEA requirement
16 by ensuring that kids with disabilities are part
17 of the accountability provisions as well, because
18 all IDEA in '97 really required was public
19 recording of data.

20 So, we have seen a lot of advances in
21 assessment. I mentioned the 1 percent alternate
22 assessment. If you stop and think about the first

1 requirement's role in assessments where in 2000,
2 it isn't -- we're only about 12 years now into
3 states' use of alternate assessments. And so I
4 would have to say that it's pretty remarkable how
5 far we've come in a very short period of time.

6 Let me show you. Here's an example of
7 alternate assessment data that happens to be from
8 New Mexico, where we can actually show growth
9 data. I mean, there was a point not that long
10 ago, a few years ago, where people said there's no
11 way we can ever do -- we would ever -- could ever
12 implement growth models on alternate assessments.
13 Well, not true. We're now doing it just 12 years
14 into this requirement.

15 The advances in accommodations,
16 universal design principles, down to assessment, I
17 know there's -- in the Disability Committee, among
18 disability advocates, there still is a lot of
19 concern about adaptive assessment. My own view is
20 that it helps us deal with something as --
21 everything is a very big challenge for many kids
22 with disabilities that are in the regular

1 assessment, and that is that when they sit down in
2 front of an assessment they may not be able to do
3 the first item they encounter. And of course for
4 lots of kids with disabilities what happens -- and
5 probably kids without disabilities -- what happens
6 is that many of them shut down. And so I don't
7 think we maximize the ability of kids to
8 demonstrate what they can do when they start an
9 assessment and they can't attempt the first item.
10 So, the use of an adaptive assessment is -- and
11 advances in adaptive assessment, I think, is a
12 very important conditioning.

13 The other thing, of course, that Alexa
14 demonstrated in her slides is the degree to which
15 we've been able to include a virtually higher
16 percent of students with disabilities in our state
17 assessment systems and, in my view, represents a
18 tremendous advancement.

19 We have some continuing issues on
20 assessment. One very big issue, and this does not
21 apply just to disability, is kind of a one-time
22 aspect of our large-scale assessments. I once

1 heard someone say that the use of our general
2 assessments, given at the end of the year is like,
3 you know, the coroner pronouncing the body dead,
4 you know, and there's no opportunity for teachers
5 to really intervene with students when, at the end
6 of the school year, the assessment happens there.
7 So, one of the advances that is being implemented
8 in many, many schools, districts, and states
9 around the country is the use of continuous
10 progress monitoring, which is often in the context
11 of RTI, which can provide teachers ongoing
12 feedback on performance of students. I know some
13 states have also implemented benchmark assessment
14 programs, which can also do the same, but my
15 observation of that is that often benchmark
16 assessment systems are not well used by teachers
17 and/or effectively used by teachers.

18 Another issue is -- a continuation, in
19 my view, is the use of portfolios for alternate
20 assessments. I think that in 2000, with states
21 having very limited lead time to implement
22 alternate assessments and the lack of any real R&D

1 that had previously existed for alternate
2 assessments, it was not surprising that states
3 would have implemented portfolios. I think -- my
4 own view is that time has come to eliminate, at
5 least for accountability purposes, the use of
6 portfolios. And, fortunately, I think there are
7 lots of advances in our assessment, and it's now
8 possible to do that.

9 Another issue I think is consistent
10 accommodation policies across states. There are a
11 number of accommodations that continue to be quite
12 controversial. One of the most is the use of
13 read- aloud in reading. Some states do permit
14 that; some states -- more states, of course, do
15 not. I think it will be one of the things -- as
16 we move toward approaches of states using common
17 assessments, I think this will be one of the very
18 big issues for us to deal with.

19 And I might say that I think that one of
20 the issues, at least in the issue of accessible
21 reading assessments is the need for innovation and
22 advancements in how we assess kids that really

1 have intractable reading disabilities. In some
2 ways for some children it's no different than
3 being blind, and so I think it's an issue that we
4 really do need to come to grips with.

5 Out-of-level testing has been an issue.
6 I think at one time when I was in the department,
7 it appeared to me that it was resolved, I think,
8 in the context of adaptive testing. It's an issue
9 again. My own view is that it's not necessary to
10 use out-of-level items on adaptive testing to
11 provide access to grade-level content, or at least
12 grade-level questions, for students with
13 disabilities.

14 Another issue in adaptive testing is
15 assessing higher-order content. I think that --
16 unfortunately, I think, in general, less difficult
17 items often tend to measure lower-level outcomes,
18 and that's another example where that doesn't need
19 to be the case; that is, we can measure problem
20 solving and higher-level outcomes with easier test
21 items. And so -- but I think -- unfortunately, I
22 think that's not typically the way items are

1 constructed.

2 In addition to issues around assessment,
3 there are many issues also associated with
4 accountability, and so currently one very big
5 issue is they should be end-sized. At one time
6 states were permitted to have different end sizes
7 for different subgroups. Fortunately, that is no
8 longer possible, because, oddly, some of the
9 smallest populations had some of the largest end
10 sizes in some cases. But we have a circumstance
11 now where there's incredible variability from
12 state to state going from an end size of five in
13 one state to as many as a hundred in other states.
14 And this is a very big issue, and it's one of the
15 primary issues in many states why sometimes even a
16 majority of kids with disabilities, schools in the
17 state are not -- I mean, for -- the majority of
18 kids with disabilities in the state are not in
19 schools where their school is accountable for the
20 subgroup.

21 Standard setting, again, I think is a
22 big issue. It relates to this fixed and -- one of

1 the reasons why I believe that the alternate
2 assessment against modified achievement standards
3 -- I mean, the modified is in front of achievement
4 standards and not in front of assessment, and I
5 think a lot of the work that went on in states had
6 to do with, you know, how do we modify assessments
7 and that the alternate assessment was really about
8 modified standards. And in some ways I think it's
9 not surprising that there may have been a need for
10 modified standards, because there were large
11 numbers of kids with disabilities who really
12 didn't belong on the alternate against alternate
13 standards. That is, they didn't have cognitive
14 disabilities. But they nevertheless were markedly
15 below proficiency on the standard assessment, and
16 this of course, absent adaptive testing, created
17 huge issues in terms of the assessment being
18 accessible, but it also included huge issues with
19 regard to the likelihood of kids being proficient.
20 And so I think that that issue, even with adaptive
21 testing, doesn't -- the issue about the large
22 number of kids being markedly below proficient

1 doesn't go away. So, there is an issue, then,
2 with the kind of having an affixed standard for
3 proficiency or moving to something that would be
4 kind of a progress standard.

5 But the progress standard doesn't solve
6 the problem, because the issue then is how much
7 progress do we expect? Do we expect for very low
8 achieving kids who may not be gaining anything
9 close to a year's -- making a year's progress in a
10 year's time? Is it likely that we're going to
11 have progress standards for kids that would be
12 less than a year's growth based on a year's
13 schooling? And so the issue doesn't go away, and
14 I see it as being one of the very challenging
15 issues that we still have to come to grips with,
16 given the tremendous heterogeneity of the students
17 with disabilities population.

18 And I think the bottom line in all of
19 this is that as challenging as some of the
20 assessment issues are, I think that we're in a way
21 better place on assessing kids with disabilities
22 than we are in instructing kids with disabilities,

1 and that is that I think even though I see myself
2 as an assessment person, oddly, my feeling is our
3 biggest issues with regard to assessing kids with
4 disabilities is not really -- let's say,
5 accountability for such disabilities isn't really
6 the assessment issues; it's really the instruction
7 issues. And in some ways I feel -- and this is
8 based even on some work I've done where when the
9 results are poor in the assessment, the belief is
10 that there's a problem with the assessment. And,
11 frankly, in the work that I've done, that's often
12 just not the case. It's not that we can't improve
13 assessments. I think we can. I think, you know,
14 we need to continue to do that. But I think that
15 it can serve as a distraction away from -- the
16 critical need is we have to improve instruction
17 for kids with disabilities.

18 Now, it's important, of course, to
19 maintain challenge and expectations for each
20 student with a disability that in my view means
21 that one notion of -- a single notion of challenge
22 and expectation can apply to all kids with

1 disabilities. We need to, in addition, create an
2 accountability system that incentivizes
3 improvement instruction. And I suppose people
4 would argue that NCLB did that. I think that
5 there are still issues.

6 If you look at NAEP data well, it's
7 clearly improved. There's still a very large
8 percentage of kids with disabilities that are not
9 proficient on the NAEP. You can see over time on
10 the upper trend line that Alexa talked about, but
11 if you just look at the percentage of kids that
12 are not proficient, this is kind of, in my view,
13 kind of -- that makes the point about the need to
14 improve instruction.

15 So, kind of looking ahead we need
16 improved assessments. We need to be looking for
17 alternative forms of accountability. And, most
18 importantly, we need to improve instruction.

19 So, with that -- I don't know how we did
20 on time.

21 MS. WEIGERT: You're ahead of the game.

22 MR. DANIELSON: Oh, good.

1 MS. WEIGERT: Now, I'd like to open up
2 questions to people at the table.

3 MS. BOWMAN: What are --

4 MS. WEIGERT: Any --

5 MS. BOWMAN: Oh, I'm sorry. What are
6 some of the models you've seen in your work that
7 bridge the gap between a student's performance on
8 the assessment and what's happening in
9 instruction?

10 MR. DANIELSON: Well, that's a good
11 question. I think -- as I articulated before, I
12 really think the notion of continuous progress
13 monitoring, as evidenced in RTI, is kind of the
14 best example in part because it provides not just
15 instructional feedback to teachers, but it
16 provides a broader system for making decisions
17 with that data to kind of intensify instruction,
18 which for kids with disabilities is kind of the
19 critical issue. If they're not making progress,
20 in the RTI framework the notion is you need to
21 then intensify instruction in some way. So, RTI
22 provides that framework, I think, for doing that.

1 MR. WIENER: Thank you, Lou. Comments
2 right on target as always, and I find them to be
3 some of the things that we're thinking about as
4 well, but you kind of put them in a nice frame for
5 us to consider as we move forward.

6 I did have a question about the comment
7 you made about portfolio assessments. Wondered
8 why the call to abandon them. Why not a call to
9 fix what's wrong with them and to make them more
10 standardized, less free-form, less driven by
11 teachers, more highly validated by experts in
12 terms of the outcomes, et cetera? Why just give
13 up on them?

14 MR. DANIELSON: Well, I mean, it may be
15 possible if what you mean, Dan, is to move to
16 portfolios that include performance events or
17 something that -- I mean, one of the critical
18 issues with portfolios, I think, is not in the way
19 we evaluate them, because I think many states have
20 gone to great lengths to develop objective,
21 independent evaluations portfolios. I think the
22 -- I believe -- I'm not going to say this is

1 universally true, but I've seen so many examples
2 where the problem with portfolios is whether it
3 reflects the students' work. And, to me, I've
4 heard people compare them to science projects and
5 how many science projects were ever solely the
6 work of the student. I know every one that my
7 daughters did I was deeply involved with.

8 (Laughter) And I think that that's
9 the challenge, and there are some
10 -- so I don't want to diminish that
11 there -- you could probably point
12 to some examples where that problem
13 has been solved, but I have seen,
14 personally, lots of examples where
15 the proficiency rates are so high
16 on these assessments. So, it's
17 either -- I mean, it's either a
18 problem where the proficiency
19 levels have been set way too low
20 and we're not expecting enough or
21 -- and, frankly, I don't think that
22 that's it. I think the problem is

1 one of who's work is it: Is it the
2 work of the student or not? And so
3 I have grave concerns.

4 And, like I said, I think that at one
5 time we had no choice, but I think now there are
6 other options. I think that we'd be better
7 served, in my view, if we looked at these other
8 options.

9 MS. FEDORCHAK: Lou, again, thank you
10 for your comments. You mentioned something about
11 grade level and being able to assess kids with
12 disabilities, kids with -- any child with any need
13 at grade level. And I fully agree with you. I
14 think that there is so much more that we can do
15 through accessibility tools that we haven't begun
16 to really tap yet, that we're just at the
17 threshold of learning how to tap. Couldn't agree
18 more with that. But I guess I'd like to ask you
19 to expand more on that comment, because I guess
20 there's a part of me, you know, that's sitting
21 here thinking, you know, really the problem isn't
22 with grade levels. The problem is grade levels.

1 You know, we don't have enough progressions. We
2 have too many grade levels and too many
3 time-stamped places to be cognitively -- we need
4 more progressions so that we can look at this
5 progress monitoring. Could you talk a little bit
6 more about your thoughts about that tension
7 between grade level progress monitoring and all
8 that that entails?

9 MR. DANIELSON: Well, as I've thought
10 about this issue, I thought about kind of the
11 multiple purposes of assessment, and I think in an
12 accountability system it is important to report
13 among, for example, fourth graders, the
14 performance -- as we hold states, districts, and
15 schools accountable, attempt to hold schools
16 accountable then for all fourth graders against
17 some common grade-level expectation. I think
18 there is a role, though, for assessments that you
19 might call out of level as we do diagnostic
20 assessment or assessments that are designed to
21 individually assess students that might be perhaps
22 for instructional purposes, where if the student,

1 frankly, in a skill area is functioning well below
2 grade level, there's a role for assessments that
3 assess students based on where they're
4 functioning. But I think that's a different
5 purpose for assessment than our large-scale
6 assessments we use for accountability purposes.

7 MR. ROONEY: Joe, can I just -- I want
8 to remind everyone at the table, since there is a
9 WebX conference call, that people are listening
10 in. If you could identify yourselves before you
11 ask your question, that would be helpful for
12 everyone to know who's talking. Thanks.

13 MR. WILLHOFT: That was Patrick Rooney.

14 (Laughter)

15 MR. ROONEY: Touché.

16 MR. WILLHOFT: And this is Joe Willhoft,
17 Lou.

18 In a slide that showed some growth over
19 time, I think, yes, on an alternate assessment.
20 As you know, these new accountability programs
21 place a lot of emphasis on growth, which is
22 probably a significant difference from previous No

1 Child Left Behind assessments. Can you comment on
2 any concerns you might have or concerns you might
3 not have with regard to the scale integrity of
4 these assessments as they assess students who are
5 particularly low performing, we'll say, and
6 whether or not we can trust or have a sense of
7 confidence in the amount of growth that might be
8 observed from year to year or perhaps the lack of
9 growth that might be observed from year to year?

10 MR. DANIELSON: And so I assume, Joe,
11 that you're talking, then, on a child-by-child
12 basis, yeah. Well, I think it -- my sense is --
13 and I'm not a psychometrician, but my sense is
14 that one of the challenges for individuals'
15 decision-making is that the standard errors can be
16 quite large, and so making judgments about growth,
17 given that, can be very often an issue on making
18 kid-by-kid determinations. And so I suspect that
19 it kind of, in general, may be an issue but I
20 think for some subpopulations of kids might be a
21 bigger issue. And I'm not sure that I have an
22 answer to that. (Laughter)

1 MR. WILLHOFT: Can I be afforded a
2 follow- up?

3 You also had a comment that you felt
4 that one wouldn't necessarily need to go out of
5 grade level with regard to the kinds of content
6 the students were assessed on, even if they were
7 lower performing. I think that's the case in many
8 assessments that can span multiple standard
9 deviations within grade-level content. But on a
10 scale, on a vertical scale, those youngsters will
11 have a number on that scale that probably lands
12 somewhere in a lower grade level. So, how do we
13 make sense of that?

14 MR. DANIELSON: I'm not sure this is
15 true of all states, by my understanding is that
16 most state assessments -- maybe all state
17 assessments -- already naturally have some overlap
18 so that it's not -- so, it may be kind of a myth
19 to say that a fourth-grade assessment is only
20 fourth-grade level and that it extends both at the
21 upper end and the lower end beyond that content.
22 And I guess my point was that I didn't feel that

1 it was necessary to deliberately add items to our
2 grade-level assessments to increase the proportion
3 of items coming from lower grade levels in order
4 to adequately assess kids on an adaptive --
5 grade-level adaptive assessment, because I think
6 that a decision issue where there are lots of
7 concerns -- and it's one of the concerns, I think,
8 that the disability community has is about adding
9 more and more items from lower-level content.

10 MS. QUENEMOEN: I think all four of the
11 consortia developing the academic tests are just
12 on a major clip right now to do as high quality a
13 summit of assessment for purposes of
14 accountability as we can. And we joke about the
15 timeline being too short it turns out. It looked
16 like it was so long to really do innovation, and
17 we're going to do the best work we can, and I
18 think this discussion will really help us.

19 I've learned, though, in the last 15
20 years to be cautious about the assessment approach
21 wars. There used to be math wars and reading
22 wars, and I realized that are favorites for

1 everything. And I think one of the really
2 intended and positive consequences of all of our
3 work is that we're also learning things about how
4 to get additional data from within the year around
5 these really rich Common Core State Standards and
6 the processes as well as the knowledge and skills
7 that are in them. So, I'm not ready to put aside
8 any -- and we all know we all have biases, me
9 included, in one direction or another in terms of
10 understanding how to equip teachers and schools
11 and districts with information across the year
12 that augments what is necessarily a snapshot in
13 time. The beauty of a standards-based reform is
14 that at some point you look at how kids are doing
15 against the standards. But you need a lot more.
16 So, it may be what we've been learning about
17 portfolio assessment, for example, in the last
18 many years will really give us rich and enduring
19 information about -- that will allow teachers to
20 learn how to do it better on the fly so that at
21 the end of the year the numbers keep getting good.
22 But I'm going to keep everything in my bag of

1 tricks as we work within your materials and then
2 do the best job we can right now with the summit
3 of assessment. So, just a caveat kind on our
4 discussion today.

5 MR. DANIELSON: Yeah, just to clarify,
6 I'm not objecting to the use of portfolios as a
7 source of information. I think student work is
8 important to compile to look at and to use. I
9 think there are inherent problems with it, though,
10 when you use it for accountability -- high stakes
11 accountability -- because I think that it's -- I
12 think it's very -- it's too tempting, I think, for
13 creating a dynamic for the teachers to help the
14 kids lots with their work when it's used for
15 accountability purposes. But for other purposes,
16 other than high-stakes accountability, I think
17 there's a role for a body of evidence or
18 portfolios.

19 MR. HOCK: Yes, okay.

20 MS. WEIGERT: Hi.

21 MR. HOCK: This is great. I've called
22 myself "mayhem man" this morning, because I came

1 out of the hotel and hadn't zipped my suitcase
2 shut and all my clothes fell on the -- and I'm
3 breaking the technology.

4 Isn't that the way? I was interested
5 in, you know, your slide of issues. And I'm happy
6 to report that one of the goals of the Smarter
7 Balanced Assessment Consortium is to come up with
8 a single accommodation policy for the entire
9 consortium. And when we started this I wasn't
10 sure it was possible but about a month or so ago
11 we brought representatives from all the Smarter
12 Balanced states together and kind of discussed
13 areas of tension that we felt might be there, and
14 we were just really surprised and pleased that
15 there was such uniformity around the ideas. And
16 although there are differences in current policies
17 across states now, everybody was ready to do
18 something different. So, I think we're going to
19 be ready to achieve this goal.

20 But getting back to your comments, the
21 one place where there still is some tension is
22 about whether or not you can read the reading

1 problems to students. And then, by the way, I
2 recall a phone conversation I had with you a
3 number of years ago where you were helping, I
4 think, D.C. or somebody --

5 MR. DANIELSON: Yeah.

6 MR. HOCK: -- try to resolve that issue.
7 So, I'm just interested in, if you know, if your
8 -- what your feelings are about that, and, you
9 know, one of the things we saw from our
10 conversations with the states is that it's not an
11 either/or thing but there may be the possibility
12 that you'll not read aloud at certain grade
13 levels, like higher grade levels where the
14 standards don't cover decoding anymore or with
15 only with certain students. So, I just -- how are
16 your feelings involved around that? Is it a
17 really bad idea? Or is it impossible to do it, to
18 read the reading test?

19 MR. DANIELSON: Well, you know, it's a
20 tough issue. I can tell you where I am personally
21 on this. I mean, personally I feel that when
22 reading reaches the point where it's about

1 comprehension for kids who still can't decode, in
2 effect it is a decoding test for them, and the
3 only way that we have available to us that I'm
4 aware of to assess the kid where the decoding is
5 not a part of the problem is the read- aloud.
6 Now, I currently have an IES grant on this issue,
7 and we're actually looking at work that a guy
8 named Jack Fletcher did where Jack found a pretty
9 powerful effect by only reading aloud proper
10 nouns, which on the face of it sounds like it
11 doesn't seem like much, why is it such a powerful
12 effect that he got? And as Jack explained when he
13 wrote it up, proper nouns are often unfamiliar to
14 kids and they are often irregular in their
15 pronunciation, and so kids spend lots of time on
16 them, and of course -- and if you have words they
17 spend lots of time on, then, you know, their
18 fluency goes down and their comprehension follows.
19 And so we're looking at Jack's work, but one of
20 the things Jack didn't do in his research is he
21 didn't compare it to read-aloud. So, we don't
22 know -- we know that doing proper nouns is better

1 than doing nothing, that is, providing no
2 read-aloud; but we don't know does reading only
3 proper nouns still fall short of a full
4 read-aloud, and we'll know that from the work
5 we're doing.

6 One other point, though, on this issue
7 that's important is that in the '04 amendments to
8 IDEA -- since the '04 amendments -- there is now a
9 requirement for states to provide accessible
10 instructional materials to kids with print
11 disabilities. And print disabilities is not just
12 kids who are blind; it's kids who are dyslexic.
13 And how that requirement is playing out is it's
14 meaning that these kids with print disabilities
15 are getting digital textbooks. And so more and
16 more it will be the case that we're going to
17 provide instructional access for kids who are
18 dyslexic with digital textbooks, which could mean
19 that kids get their content with a full
20 read-aloud, or it could mean that through the
21 beauty of the software that now exists the reading
22 is just supported by the computer. That is, they

1 can click on words; there are things that they can
2 do to support their reading. And with more and
3 more kids receiving instruction that way, my view
4 is that if assessment doesn't align with that,
5 we're going to have an issue. So, I think that
6 this -- I still think that no matter -- even if
7 states agree on this issue, I think that it's
8 going to continue to be a difficult issue.

9 And I don't disagree. You know, in the
10 earlier grades where the thing being assessed is
11 decoding, clearly to not read aloud in that case
12 makes perfect sense, because that's the purpose of
13 the assessment then. But I think -- as kids get
14 older I do think that, particularly for kids who
15 daily are getting their textbooks provided
16 digitally, it just doesn't make sense to me that
17 we would not assess them in a similar way.

18 MR. HOCK: I have just a one-sentence,
19 quick follow-up.

20 You know, I'd just like to point out
21 that one of the other tensions that came up in our
22 meeting with the states was sort of the

1 intersection between accommodation and the notion
2 of career college readiness and if we need to
3 think about accommodations in terms of what kids
4 would use in college, and so I think that your
5 example of digital textbooks is really something
6 we have to look at.

7 MS. LAZARUS: I'm really excited that
8 we're holding this meeting today, because there
9 are so many important things that need to be
10 discussed, and I guess Lou talked about the NCEO
11 had been around 21, 22 years at this point, and I
12 sat here and I realized, wow, on and off I've been
13 a part of this for 10 years now, and it's awesome
14 seeing the advances, how students with
15 disabilities are included both in instruction and
16 in assessment in ways that we didn't think was
17 possible years ago. But much of our work over
18 that time period has been dealing with
19 retrofitting assessments, trying to make things
20 that weren't really very well thought out about
21 how to make them accessible to make them somehow
22 accessible. And it's good that we're all here at

1 the table thinking, and we have an incredible
2 opportunity to get them right from the start. And
3 we don't want to lose that opportunity.

4 I've been asked today to talk a little
5 bit about how to incorporate some of the universal
6 design principles into the next generation of
7 assessments. Now, much of what I'm going to say
8 today is not new. I think pretty much everyone
9 here in this room probably knows everything. Many
10 of you could probably give the presentation I'm
11 going to give or do it even better than what I
12 can. But I think it's a really good, useful
13 reminder as we move forward not to lose sight of
14 those principles of universal design.

15 Now, just as a very brief refresher,
16 universal design first became a term that was used
17 back in the 1990s, and often there was a
18 discussion about how the design of products and
19 environments need to be usable by all people, or
20 at least as many people as possible, without the
21 need for adaptation or a specialized design and
22 that, really, a way to do that is to think right

1 from the start about how that might be done. And
2 examples were often given out of architecture or
3 tool design, examples about how today to even
4 think twice about whether or not cutouts and
5 curves are a good thing. They're useful. They
6 make the sidewalks much more usable for a much
7 broader group of people, not only wheelchairs but
8 also parents with those strollers.

9 Another example of universal design
10 coming out of architecture is elevators that can
11 talk to you; door handles rather than knobs; and
12 even things as very, very basic is that some pen
13 shapes are just a lot easier to hold, to use, and
14 that a pen that is more universally designed can
15 be used by a larger group of the population than
16 something that is less universally designed.

17 Now, a number of organizations have been
18 very involved with and have done a lot of just
19 absolutely awesome work on universal design for
20 learning. I'm going to give a little bit of some
21 definitions, some information on UDL coming out of
22 the work of the National Center on Universal

1 Design for Learning, though there are other
2 organizations that have very similar information.
3 But the definition is it's a set of principles for
4 curriculum development to give all individuals an
5 equal opportunity to learn. And it can really
6 provide a blueprint for that learning by creating
7 instructional goals, methods, materials, and
8 assessments that are for everyone to enable the
9 broadest group to have access.

10 The components. Let's talk a little bit
11 more about those components of UDL. There's the
12 goals, those learning expectations, and they
13 really represent that knowledge, concepts, skills
14 that all students need to master. They provide
15 information about what kinds of pathways different
16 kids might take. And the pathways might not all
17 be the same for every student. But those pathways
18 will get them all to the same place in the end. A
19 variety of methods might be used to get those kids
20 to different places. The instructional decisions,
21 the approaches, what kind of scaffolding might be
22 needed? How is this all pulled together so that

1 high expectations are not lost? Materials will
2 vary. Different kids might need different types
3 of materials, tools to get there.

4 And then the last component of universal
5 design for learning is the assessments, that
6 process of gathering information about a learner's
7 performance using a variety of methods.

8 Now, many organizations, as I've said
9 before, they've really developed some just awesome
10 tools. I want highlight one that I recently
11 became aware of that the Maryland Department of
12 Education developed using some federal funding.
13 Trinell probably knows more about it than I do.
14 But I found it incredibly exciting, and I just
15 want to highlight this, but totally recognize that
16 there are so many other tools out there that
17 really can enable teachers, educators to make
18 better use of and develop instruction that is much
19 more universally designed.

20 This Maryland tool, one of the things
21 that I like about it is that it's developed in a
22 way that educators at all places on their personal

1 scale of learning how to use technology can use.
2 So, it's in some ways universally designed for
3 those educators, that there is a hard copy UDL we
4 offer that educator that's still not quite with it
5 -- with the technology, that they can just use
6 that. There is a second form, which is the online
7 version, which hopefully we're going to show in
8 just a moment.

9 Let's see if we can get it up. This is
10 always the nervous part here with -- okay.

11 So, this is -- it's on the Internet, and
12 while they're getting that up, there's also a
13 third version of this, which is the phone app, and
14 I am just, like -- I'm in awe that they have a UDL
15 phone app.

16 And there it is. Now, the wheel is
17 interactive. It can be turned to show different
18 things. The first one is options for perception,
19 and then the different sections on the wheel,
20 which then as just was here a minute ago, shows
21 different things that can be used -- strategies
22 that an educator might use, for example with the

1 options for -- well, it just went away. But for
2 self-regulation, they might use props, checklists,
3 those kinds of things. There it is. Use,
4 suggestions that an educator might use. Using
5 prompts, reminders, guides, rubrics, checklists,
6 so on for self-regulation, some other things.

7 Designing activities, which the learners can get
8 feedback and have access to alternative scaffolds.

9 So, anyhow, that's just one really fun
10 thing that is available. Not only fun but really
11 useful, and so let's go back to the main slides
12 now.

13 Okay, now, talked a little bit about
14 UDL, some about instruction. And I absolutely
15 totally agree with Lou that instruction is key,
16 that we're here, we're all talking about
17 assessment, but the issue is really instruction,
18 and assessments really provide that window and
19 then can flag when there are instructional issues
20 and how can we improve instruction. But I'm going
21 to focus more at this point on universally
22 designed assessments. And universally designed

1 assessments are designed right from the beginning
2 to be accessible and valid for the widest range of
3 students and provide optimal standard assessment
4 conditions that in some cases, accommodations may
5 need to be used but to minimize that need, because
6 the assessments have been well designed, well
7 thought out right from the start.

8 Universally designed assessments, those
9 tests remain true to the constructs, are easy to
10 understand, and contain language that is
11 accessible to all and will give the truest
12 readings of what students do and do not know. And
13 universally designed assessments are -- which
14 sometimes I hear, but it's just not true, that,
15 you know, they are not dumbing down the test, that
16 they are remaining true to the constructs.

17 Who benefits from universally designed
18 assessments? It's not just kids with
19 disabilities, English-language learners. All
20 students benefit from better-designed tests, have
21 better opportunities to access what the test is
22 answering, and respond appropriately. And it

1 applies to all students.

2 Now, I'm going to talk just a bit about
3 two GSEG projects that I've been very involved in.
4 We've done a little bit of work related to
5 universal design.

6 The first is the multistate GSEG
7 consortia, and this is a consortia of five states
8 -- Alabama, Hawaii, South Dakota, Tennessee, and
9 Wisconsin -- that wanted to look at: What were
10 the characteristics of persistently low-performing
11 students? And then what are the issues? Is it an
12 instructional issue? Is it an assessment issue?
13 How can we better instruct and assess that group
14 of kids?

15 The consortia decided early on that they
16 could do more as a group than what any one
17 individual state could do and that they all should
18 not do exactly the same thing, that by exploring
19 different options they could share with one
20 another, identifying different issues; that for
21 accommodations, three of the states in the
22 consortia looked at issues related to

1 accommodations, and I'm trying to -- let's see --
2 and that was Alabama, South Dakota, and Tennessee;
3 that two states looked at the nature of
4 instruction that was received -- South Dakota,
5 Tennessee, and Wisconsin; and then Hawaii did some
6 really awesome work related to learning
7 progressions.

8 Lou's talked quite a bit about the read-
9 aloud accommodation in reading. One of the things
10 that this consortia, or at least one state in the
11 consortia, did is looked at some issues related to
12 the read-aloud accommodation in math, and that's a
13 widely used accommodation, and we really don't
14 know a whole about what is happening with that
15 accommodation. South Dakota has looked at that
16 accommodation as very broadly used there and first
17 did some focus groups with test administrators to
18 see, from their perspective, what's happening,
19 what's working, what's not working, what are the
20 issues.

21 And then we are still in the process of
22 completing some think-alouds to really get the

1 student perspectives. We're administering to each
2 student ten items, five of which are read aloud to
3 the student and then five of which the student
4 reads him- or herself, and the items are paired,
5 but they're very, you know, similar-type items, to
6 really get a sense of are these kids -- are they
7 doing better when they have the read- aloud
8 accommodation in math? Not better? About the
9 same? What are the characteristics of the kids
10 that it's helping? And so we're really looking
11 forward to getting those results very, very soon.

12 I'm going to now sort of step back from
13 the detailed level, but I'd be really happy to
14 discuss -- and if you have any questions I can
15 tell you the details of many of the other studies,
16 because there's been just a lot of awesome work.

17 But the final activity of the multistate
18 GSEG is something that I think will be of a lot of
19 interest to the consortia, and that is that we are
20 publishing a report on the lessons learned by this
21 group of GSEG consortia, as well as EIG grants and
22 some supplemental funding grants that looked at

1 issues related to low-performing students with
2 disabilities, invited the PIs of the different
3 projects to write a chapter. I believe fourteen
4 of the PIs took us up on it, so it will be just an
5 awesome resource that is coming out later this
6 summer, early this fall where it will be kind of a
7 one-stop place to just really get some of this
8 information so that it's not lost.

9 Now, a very, very closely related
10 project that I've been involved with is the
11 Alabama GSEG. And the Alabama GSEG has done a
12 number of things, but one of the things that they
13 really cared about was universal design. And they
14 had a lot of questions about what universal design
15 meant and what it meant in relationship to
16 low-performing students with disabilities. So, we
17 pulled together a panel of experts and kind of
18 picked their brains on some of these issues.

19 Some of the questions that Alabama had
20 were: What strategies can improve the universal
21 design of assessments for low-performing students
22 with disabilities? And then maybe it's the

1 elephant in the room -- it's a thing that we all
2 worry about, it's a thing we don't talk about a
3 lot -- but just what is universal design versus
4 what is a lowered-achievement expectation, and
5 trying to get an understanding of that difference?
6 So, Alabama started looking into the literature,
7 just what it says, and found some things that were
8 very, very useful. And let's look at a few of
9 them.

10 Now, the elements of universally
11 designed assessments, and this is pulling from the
12 work that was done a few years ago by Sandy
13 Thompson, Martha Thurlow, and Chris Johnstone.
14 And they identified seven elements of universally
15 designed assessments, that it -- and a universally
16 designed assessment has an inclusive assessment
17 population that for the target population most
18 students are able to participate.

19 It's important to have precisely defined
20 constructs. What is the assessment designed to
21 measure? And having an understanding of what the
22 assessment is designed to measure really helps

1 think through what is an appropriate features
2 element that should be included in making that
3 assessment more accessible.

4 The third characteristic or third
5 element of universally designed assessments is
6 that it is important that they have accessible
7 non-biased items, that some assessments develop
8 items that really just don't work for certain
9 populations, be it students of a certain
10 ethnicity, socioeconomic group, or students with
11 disabilities or English-language learners. And
12 it's so important to, really, develop assessments
13 that are accessible regardless of the population
14 that the student comes from.

15 A fourth characteristic is that it's
16 amenable to accommodations. Now, as I previously
17 mentioned with assessments that are more
18 universally designed right from the start, there
19 will be less need for accommodations. But there
20 always will be a need for accommodations, but
21 we'll never have the absolutely perfect assessment
22 that meets every student's need.

1 Universal designs on a continuum.

2 Hopefully, we're getting over a well thought out
3 assessment with one that's much more universally
4 designed, but there still will be a need for
5 accommodations to enable some students to
6 meaningfully access that test. But thinking
7 through, right from the start, the assessment that
8 is designed in a way that accommodations can be
9 used easily with that assessment for all of the
10 populations that might be taking that assessment
11 is absolutely vital.

12 The next element of universally designed
13 assessments is that it has simple, clear,
14 intuitive instructions and procedures, that the
15 instructions should be written in a way that both
16 the administrators as well as the students taking
17 the test understand what they're supposed to do,
18 that if the student gets lost or sometimes the
19 administrator sometimes gets lost in the
20 instructions and it's not clear what needs to be
21 done, you've never even gotten to the assessment
22 itself. So, vital that clear instructions -- and

1 then -- and I know we're moving toward
2 computer-based tests.

3 The new assessments will be
4 computer-based tests. But the last two are still
5 important, even in that new realm, and that is
6 that it's maximum readability and
7 comprehensibility, that it's thinking about how
8 the information is organized, structured so that
9 it's easy to read -- short sentences, clear
10 sentences -- just making it as readable as
11 possible.

12 And then maximum legibility -- and even
13 within the framework of a computer-based test,
14 some things can be more legible; other things,
15 less legible -- and really thinking that through
16 as the design right from the very beginning.

17 Now, I gave some elements of universally
18 designed assessments. Now, let me give you
19 another group who has some universal design
20 characteristics. And now I'm moving down in grade
21 size from the test itself to the individual items
22 and thinking about universal design at the item

1 level.

2 Now, individual items on a universally
3 designed assessment. Each item should measure
4 what it intends to measure, that it should respect
5 the diversity of the assessment population. It's
6 really important that there be a clear format for
7 the text and that there are clear pictures and
8 graphics when essential for the items. And
9 thinking it through, even within the world of --
10 there are so many absolutely cool things that can
11 be done with computer- based tests but still
12 ensuring that it's done in a way that it really
13 doesn't muddy the waters but is helpful for the
14 items. Important to have clear and readable text,
15 as well as allowing for changes in the format
16 without changing meaning more difficulty.

17 Now, I'm not going to go into much
18 detail, but it's so important to really think
19 about what that layout and design would be on
20 computer-based tests, thinking about doing it in a
21 way that the layout -- that all the information
22 needed whenever possible to answer a given item is

1 on a single screen, maintaining consistency in
2 format as much as possible from one item to the
3 next because the format's totally changed from one
4 to the next, it can be very confusing; thinking
5 about the navigation and how to do that in a way
6 that is intuitive and not confusing, simple to
7 understand; and then also considering what the
8 computer capabilities are and ensuring that the
9 assistive technologies that some students may need
10 to access the assessment that they are there.

11 Now, I have a couple of recommendations.
12 First is to, as we said before, start right at the
13 beginning. And then it's very important to
14 include disability and language acquisition
15 experts in those item reviews, people that really
16 have the knowledge as to how students with
17 different disabilities access those items. It's
18 important to provide professional development and
19 to present the items being reviewed in a format
20 that will appear on the test, as well as including
21 the standards being tested with the items being
22 reviewed. And it's vital to try out the items

1 with the students using think-alouds and really
2 getting to know what works, what doesn't work, as
3 well as field testing.

4 This is an opportunity, and I'm going to
5 close here, but I just beg of you and the
6 consortia not to lose, to take advantage of this
7 opportunity and make the most of it, because we
8 really do have just an incredible opportunity to
9 develop tests that are accessible to all students.

10 Thank you.

11 MS. WEIGERT: Two questions for Sheryl
12 the next few minutes on universal design.

13 MS. POSNY: Can I ask one?

14 MS. WEIGERT: Sure.

15 MS. POSNY: I know this is kind of out
16 of order, but a question I have for you, because
17 it's something that I've heard from the disability
18 community, has to do with the design and layout,
19 especially on the computer. One concern or one
20 thing that's been voiced, especially by the deaf
21 and hard- of-hearing community, is that they're
22 under the impression -- and you guys dispel this

1 myth if this is true -- that, you know,
2 interpreting -- you know, sign language
3 interpreting -- that some of the assessment
4 designs are going to use avatars since it's going
5 to be on the computer. And they're very
6 concerned, because that's not a methodology that's
7 used in terms of instruction. So, what -- I mean,
8 how would you respond to that?

9 MS. LAZARUS: I would respond to that by
10 saying that I know a lot about accommodations in
11 general. There are people other than me that have
12 much more expertise in accommodations for the deaf
13 and hard of hearing. Probably some of those
14 people are individuals that have been talking with
15 you, with the consortia. It's important to listen
16 to them. They know what the issues are, what the
17 problems have been in the past, what some of the
18 concerns might be. Listen, talk to them. The
19 assessment that you develop, no matter how well
20 it's developed, is not going to be the perfect
21 assessment. But by using a transparent process,
22 talking to the people with the most expertise in

1 the different areas, it will be a much more
2 inclusive, a much better assessment than if they
3 are not all involved.

4 MR. WIENER: Hi, this is Dan from
5 Massachusetts.

6 The avatar question that PARCC is
7 grappling with or will soon grapple with, we've
8 already started discussions about whether it's
9 meritorious or not. But the bigger question for
10 us, I think, is how does universal design exist
11 when you're kind of constantly thinking about
12 different populations with mutually exclusive
13 needs? And I'm thinking about how we spend
14 endless amounts of time adapting Braille tests,
15 the graphics for Braille tests, for blind users,
16 and yet deaf students benefit from having very
17 vibrant, complex graphics because that's a way in
18 which they learn. And it makes me skeptical that
19 either the retrofitting part of this is never
20 going to go away or we need so much time to
21 prestock the background of each item so that it
22 can be given appropriately on a computer for the

1 population that we won't have enough time to do
2 that.

3 If you have worked in the assessment
4 world, you'd know that the timelines are very
5 tight, the deadlines are ridiculous. We're always
6 down to the last moment. But with APIP and other
7 exportable item formats, we now have the
8 opportunity to kind of stock the item with all
9 this background information, that at the press of
10 a button the read-aloud comes up or a sign --
11 either it's a videotape or an avatar comes up on
12 the screen or the adapted -- the text that
13 replaces the graphic comes up, or whatever it is
14 comes up.

15 But we -- somebody has to do that. And
16 so the universal part of it may be universal, it
17 looks universal to the user, but somebody's always
18 fitting and retrofitting, and so that takes an
19 incredible amount of time, and I think, you know,
20 if there was an elephant in the room, that might
21 be it. Technology might help us by reducing the
22 number of accommodations that we call

1 accommodations and others that we might call
2 embedded features that all students could use --
3 for example, enlarging the text, making a
4 background color different, and so forth.

5 But still, I'm a little bit skeptical.
6 Maybe you can respond to this. Even with
7 technology, I just don't know how you restock
8 items every year and refresh items every year and
9 put all this information in for every item on
10 every test for every kid. I mean, it really does
11 sound like pie in the sky, and -- what do you
12 think?

13 MS. LAZARUS: Now, my understanding of
14 APIP, and please correct me if I don't have it
15 quite right, but my understanding of APIP, at
16 least stepping back from the details, is that one
17 of the purposes of it is to identify items that
18 may not be appropriate to be Brailled or
19 indicating which items can be, what, making them
20 accessible in different ways, but that some items
21 may be more amenable to being made -- used in
22 Braille than others. Is that --

1 MR. WIENER: As I understand it, APIP is
2 a way to do all kinds of very wonderful and
3 interesting and creative work around creating more
4 accessible items for individual kids who have an
5 assessment profile or an accommodations profile.
6 And so all this stuff can be programmed in through
7 an exportable item format, like APIP, and at the
8 press of a button your profile comes up and you
9 get your accommodations and they've all been
10 stocked into the item so that you can take the
11 test your way. It's a great thing, and it's open
12 to human creativity. But it's also a way of
13 exporting items from one test publisher to another
14 so that the features of the item, the displays of
15 the item, remain in tact.

16 MS. LAZARUS: And that's my
17 understanding of it also, that, you know, it is to
18 really come up with -- the student profiles make
19 it more accessible.

20 The devil's in the details, that APIP
21 has some -- as well as other similar, what,
22 methodologies -- has some really wonderful goals

1 going in the directions that we really needed to
2 go for a long time. However, as you indicated,
3 Dan, it's just incredibly complicated, getting
4 everything coded, programmed the way it needs to
5 be. With APIP we're on tight timelines. It's
6 also important to get item writers and those
7 involved in developing these items up to speed or
8 ensure that computer programs are doing it
9 appropriately, and there are a lot of issues there
10 that I know are being sorted out, and it's hard.

11 MR. SHEINKER: Alan Sheinker. We are --
12 at DLM, we've developed a learning map, and what
13 we're trying to do to accommodate these individual
14 issues is look at alternate pathways that students
15 would not take normally -- visually impaired
16 students or students who are deaf, and create
17 those pathways up front so that these students,
18 through a profile about the student, would not be
19 presented items that we would then have to discard
20 or invalidate. So, we're trying to deal with it
21 up front. Then we also will have psychometric
22 issues that we'll have to deal with in

1 comparability, but I think we can do that, that
2 we're trying to meet that head on.

3 MS. WEIGERT: We have one more.

4 MR. HOCK: That one works. Question
5 from Michael Hock.

6 Alexa, when you brought up signing
7 avatars, we're also running into, you know,
8 similar questions about using Braille embossers
9 and that sort of thing. And what I thought about
10 when you asked your question is kind of the old
11 saw on accommodations has always been that we
12 should only provide accommodations that students
13 use regularly in the classroom. But now we're
14 talking about sort of the age of technology, and
15 in many cases if we apply that, what it means is
16 we would build our tests kind of around the lowest
17 common denominator; that is, we wouldn't use
18 technologies unless they're available everywhere.
19 And we know, for example, that, you know,
20 delivering tests to students who were blind,
21 visually impaired, that there are a lot of
22 challenges around that. And the Braille embossers

1 is a great thing, but our understanding is they're
2 not being used in that many classrooms yet. So,
3 you know, I guess my question for Sheryl is, you
4 know, does that old saw still apply? Or can the
5 tests really drive getting better technologies in
6 the classrooms for kids?

7 MS. LAZARUS: I think that it's what --
8 something that -- we are moving toward more
9 technology, that it's vital that we don't put a
10 student in a bind where we're saying hey, you
11 don't know how to use this piece of equipment.
12 You don't use it in the classroom, but here.
13 You've got to use it on the test. That's not
14 good. That's a problem. But it's -- but at the
15 same time, if we could get those Braille embossers
16 in the classroom, get the student using some of
17 these things with -- you know, having it used with
18 their educators, with the students, with
19 instruction, that would be a good thing. And so
20 the old rules still apply, that students shouldn't
21 suddenly encounter an accommodation on an
22 assessment that they haven't used in instruction.

1 But I guess the challenge to you is to really make
2 sure that the things that would help the students,
3 that would improve their learning are being
4 regularly utilized in instruction so that it
5 becomes a non-issue by the time it gets to
6 assessment day.

7 MS. WEIGERT: Well, thank you, Sheryl,
8 and I'm going to call for a 10- to 12-minute break
9 if we can do that and -- 10- or 12-minute break if
10 we can. Thank you.

11 (Recess)

12 MS. POSNY: Thank you. We're going to
13 start the next presentation in a moment.

14 MR. ELLIOTT: Hello, folks. Thank you.
15 My name is Steve Elliott, and this is the third
16 session, Improving Accessibility to Large-Scale
17 Assessment for All Students. I actually do want
18 to make eye contact with you. I really feel rude
19 having my back to you, but I'll do the best I can
20 to project my voice and engage you.

21 This is an exciting opportunity for me
22 as a researcher. I consider this the impact zone.

1 This is where people take ideas and, I think, put
2 them into practice; and I'm thrilled about the
3 opportunity to share work of many, many people
4 that I will not acknowledge due to time, but
5 they're documented in references, et cetera,
6 today.

7 This work fits under, I think, a rather
8 large umbrella, from my perspective, inclusive
9 assessment that allows for valid test score
10 inferences. Those are partners here. Inclusive
11 assessments? Sure, we can build assessments that
12 are inclusive, but at the end of the day they also
13 have to yield valid inferences about the test
14 scores that we're interested in.

15 So, two big ideas to push forward.
16 One's already been, I think, introduced and talked
17 quite a bit about, and that's item modification or
18 test modification or adaptive assessments, however
19 you'd like to phrase it. I want to bring some
20 detail and specificity to universal design
21 practices and principles. And, secondly, I want
22 to talk about opportunity to learn and a critical

1 role in understanding and ensuring that students
2 having actually had an opportunity to learn the
3 content they were being tested on. So, let's move
4 forward.

5 The first slide here, actually slide 2,
6 two questions for you, and I show it now and I'll
7 address these at the end. You should be able to
8 answer these questions if I've done a good job
9 today.

10 How can tests be modified to improve the
11 measurement of knowledge and skills of students
12 with persistent academic difficulties, and who
13 would qualify to take -- if they would qualify to
14 take -- formerly what we'd call an AAMAS if
15 available?

16 I want to just stress here that I chose
17 my title carefully. It was for all students.
18 What I have found in all this research that I've
19 done is, first off, we always do research with
20 students, with teachers. They're partners in
21 helping us think and understand, and we always
22 include students with and without disabilities.

1 And, as you all know, students with disabilities
2 and without disabilities are far, far more alike
3 than they are different, okay? So, it's important
4 to understand.

5 Secondly, how can testing conditions be
6 improved to increase the likelihood that all
7 students have the opportunity to learn the content
8 standards that tests are designed to measure? I
9 think we have some new breaking information on
10 this that will not necessarily be comforting that
11 I'd like to share with you, so.

12 As we go forward, it is really important
13 to acknowledge both my various partners, many
14 state partners, many of whom are in this room --
15 thank you very much -- and also university-based
16 researchers and colleagues. And without the
17 funding of the U.S. Department of Ed, basically I
18 don't think I'd be sitting here sharing this
19 research with you, because it is, in fact,
20 expensive research, time-consuming research, and
21 it takes lots of teamwork to do. So, acknowledge
22 those projects here.

1 So, the work I mentioned -- I frame it
2 as inclusive assessment, and in doing that we've
3 tried to think about how can we optimize access to
4 assessments for students. So, that's kind of the
5 main notion. Let's come up with a pathway or
6 pathways, as Alan reminds us, to overcome barriers
7 to assessment.

8 So, the visual here is what are the
9 barriers? There are probably many, but what I'm
10 going to focus on, one, is insufficient
11 opportunity to learn the intended curriculum.
12 Many people have talked about it and although I
13 don't think it's passé to think about testing
14 accommodations as a critical element, to me the
15 big challenge isn't the testing accommodations,
16 it's making sure they get delivered, although
17 today I'm not going to speak about that given the
18 other things that I have to do.

19 And then, finally, I want to talk about
20 constructor relevant variance on assessments, in
21 other words, ways to improve test items so that
22 they're measuring what we say they measuring,

1 okay? And those will be the two high points of
2 this.

3 So, jumping into the first section of
4 this, I'm sorry to be simplistic, but in terms of
5 thinking about designing items that are highly
6 accessible in the work that we've been doing, and
7 most of the items are multiple-choice items, I'd
8 like to encourage you to think about an anatomy of
9 a test item. And so I've laid out an example of
10 that so that we could have our state partners and
11 our teachers and our students help us think about
12 the pieces of a test item. And there they are:
13 Item stimulus; visual -- not items have visuals,
14 mind you, but this one does; item stem; and answer
15 choices. And so then the overall layout on a
16 page. Sounds simple. These are, I think, the
17 five elements that if we start focusing on items
18 and analyze them, we can find -- we can make
19 adjustments to one or more of these that make a
20 difference, okay?

21 I want to highlight, though -- actually,
22 I probably should go back to that slide. What I'm

1 going to be talking about are really packages of
2 changes. In our work that we've done, we've
3 changed an item. We haven't tinkered with just
4 one dimension of the item, so there often are
5 questions. Which one makes the most difference?
6 I can't answer that right now. I can tell you
7 which packages of item modifications seem to make
8 difference.

9 So, this work has been guided largely by
10 theory, but in some cases where theory is lacking
11 we have actually looked at existing research.
12 Probably the thing I want to highlight on this
13 page, there are four points. The evidence-based
14 models for test score validity, so we've always
15 worked within sort of an AERA/APA/NCME 1999
16 validity framework in thinking about evidence that
17 we need in order to determine if in fact items are
18 comparable, et cetera. I'll say more about that.
19 I imagine Steve Ferrara will say more about that
20 later today.

21 Yes, we've talked about universal design
22 principles, but in many ways there's not enough

1 specificity, although I think we heard some
2 specificity today from Sheryl as she talked about
3 it.

4 The most important and I think the new
5 piece of information that has come out of this
6 body of work is work that's been done in
7 instructional psychology under the heading of
8 cognitive load theory. And if you just think
9 about that, it largely deals with how information
10 comes to us in packages and can overload us, so
11 one of our jobs in this work is to apply -- there
12 actually are 31 principles of cognitive load
13 theory that John Sweller and colleagues have come
14 up with over the last two decades, and we've found
15 that we can apply these to the design of test
16 items and that they can, in many cases, make a
17 difference.

18 Finally, there's a lot of item writing
19 research, some folklore, too, by the way, out
20 there that we can use to guide our work. A good
21 example of item writing research, as far as I'm
22 concerned, would be the work of Michael Rodriguez,

1 and many of you know his work in answer choices,
2 et cetera, but that's just the beginning, the tip
3 of the iceberg, a side of his work that's
4 relevant.

5 The next slide, titled "Characteristics
6 of Appropriate Modifications," so, here are the
7 design elements that we went through as we, in a
8 sense, revised items. Let me just say we didn't
9 write new items from scratch, we worked with our
10 state partners. They gave us pools of items, and
11 we started looking at these elements. In effect,
12 we assessed items on these dimensions and more,
13 frankly, "simplify words and text structure,
14 delete extraneous words, improve visuals, and
15 locate" -- "location" it could be -- "within the
16 item, use bold text for important words, examine
17 answer distracters for plausibility, and also
18 disproportionately" -- it says "disproportionality
19 of selection of the item."

20 So, to accomplish those things, we
21 needed a framework and a structure, but if we do
22 those things, we think these are the results,

1 these are the desired outcomes that we were
2 looking for in our work. One -- and I'm going to
3 define accessibility in a minute, but we wanted to
4 increase the overall accessibility so that
5 virtually all test takers could be engaged. And
6 we wanted to, again, under this work, under the 2
7 percent alternate assessment, think about decrease
8 item difficulty, increase item discrimination,
9 increase reliability estimates, reduce readability
10 level, but stay within a grade range for
11 readability that the existing test utilized. We
12 wanted to maintain alignment with the content
13 standards. We wanted to maintain the depth of
14 knowledge, DOK, for all items. We want to reduce
15 the need for accommodations, not necessarily
16 eliminate, but we think we can reduce the need for
17 accommodations, and we want to reduce the number
18 of words and at the same time improve the
19 student's perceptions of tests and their
20 motivation to engage in the tests.

21 We had a lot of kids say when they look
22 at the tests they totally freak out, and so how

1 can we present -- kind of like, let's make sure
2 the first item they can engage in, and that's
3 maybe kind of a silly point, but not so, actually.
4 So, if we do this, we think the results will --
5 they increase test score validity. Now, that's a
6 big menu. A lot of desired outcomes. I don't
7 know if you share all of them. But I think -- I
8 tried to cover them all, and we have data on all
9 of those, okay?

10 So, here's an example of an item. This
11 is a real item. I'll read it for those who can't
12 see it very well from where you stand. It says,
13 "Joan earns \$100 per month working part time in a
14 music store. Look at the" -- I can't see it.
15 "Look at the pie chart that shows how Joan budgets
16 her money each month. If Joan starts" -- excuse
17 me -- "sticks" -- I can't even read it, "to her
18 budget, how much can she spend on clothes and
19 entertainment each month?" So, you can see the
20 item, you can see how it's arranged, colored pie
21 chart.

22 Here's an alternative that we came up

1 with, with the NR model, okay? I'm sure you could
2 come up with others. It's actually a fun thing to
3 try to do this into revised items.

4 So, as you see right off the top, there
5 are fewer words, a black-and white colored with --
6 gridded in case children are color blind, and
7 three answer choices versus four answer choices,
8 so.

9 How did we get there? How did we -- I
10 mean, that item and hundreds of other items. How
11 did we work through? We used a tool that grew out
12 of the CAAVES grant and then the CMADDI grant that
13 we ended up calling TAMI -- the Test Accessibility
14 and Modification Inventory -- and it's a pretty
15 simple tool. It's just -- here it is, it's a
16 couple sheets of paper.

17 Why do we need it? Because we have lots
18 of people involved in working with items. We want
19 consistency of activities. We want a documented
20 record of changes, et cetera. So, it lends
21 itself, I think, to helping you to codify and
22 standardize a process of improving the

1 accessibility of items.

2 A little sidebar that's not in my
3 slides: We offered to do this for five different
4 states, in their science items and another project
5 called Oasis. Each state submitted 500 test items
6 to us at random, and analyzing those items we
7 found that 5 percent of the items, across all of
8 the states, had a high standard of accessibility.
9 Most of those items were at our next highest level
10 of accessibility, so they weren't terrible items
11 in any way. But there's room for improvement
12 virtually on all of our items.

13 And so here's the overall structure. We
14 use a four-level structure for talking about test
15 accessibility. And due to time, I'm not going to
16 read all the levels and the details of them, but
17 they build upon the elements of that item anatomy
18 that we talked about, and there are four levels.
19 Basically, level one, inaccessible for many test
20 takers. Obviously, we don't want items for level
21 one. At the top, it's maximally accessible for
22 nearly all test takers. And so as we look at

1 items and as we train item writers -- and we've
2 done that in many states as well -- we start
3 getting some consensus about important elements of
4 items and about the process of documenting those
5 items.

6 So, you get a tool -- here it is, the
7 worksheet, basically, this record -- and on the
8 left-hand side, you can see that we're back to
9 those elements of an item, the anatomy of an item:
10 Passage, item stimulus, item stem, visuals, answer
11 choices, page layout. And we yield an overall
12 score for these items. So, we would get three or
13 four people to look at the same item, poll
14 independently, and you get consensus around these
15 elements. And it's a beginning of a communication
16 process and documentation process that I think has
17 often yielded much better items in terms of actual
18 engagement out of our students and some
19 interesting aspects in terms of outcomes, actual
20 scores.

21 Here's another example of an item,
22 surreal item. The first thing is you see that

1 it's one of several items on a sheet. This is the
2 kind we don't want to start with, Lou. Kids look
3 at this and go, holy smokes, I got all this to do?
4 So, that's -- one of the beauties of a
5 computerized assessment is we can focus on one
6 item, we can spread it out, et cetera.

7 Again, for the sake of time I don't want
8 to read through this entire item, but I want to
9 show you, in effect, illustrate the process that
10 we use.

11 So, this is an item that deals with a
12 math -- I think it's a fifth grade, or its target
13 construct is use knowledge of two-dimensional
14 shapes to solve real-world calculation problems.
15 This is a grade five item from a state assessment.
16 And we asked several people to work with this
17 item, utilizing this TAMI framework to maximize
18 accessibility.

19 Here was the winning choice that we came
20 up with. First off, I want to document some
21 characteristics -- I apologize. This is obviously
22 math domain. You can do the word count, 75; you

1 can do the readability, 5.5. On the TAMI rating,
2 this was a 1. This was a bad item. That's from
3 our perspective, from accessibility. It had some
4 significant difficulties. So, we transformed it.
5 There it is. There's the transformation of that
6 item. The domain, again, stayed in math; word
7 count, 60. We reduced it by 20 percent. And I
8 should say, I jumped into this, one of the
9 pervasive elements that identify children who were
10 qualifying in the states we worked in for
11 alternate assessments at 2 percent was reading
12 fluency. Reading fluency, regardless of their
13 disability category, was a difficulty. This also
14 speaks back to some of Lou's points about Jack
15 Fletcher's work with a noun, and any time we start
16 introducing words slows kids down. They're speed
17 bumps. A speed bump to a kid with reading fluency
18 problems is a real problem now. It just
19 exacerbates the situation.

20 The chairs are in the way, but this is
21 the exact same target construct that we identified
22 previously. This is determined by other people

1 independent of our work. The readability, 3.2.
2 So, we actually lowered the readability by a
3 couple grades. Now, somebody might, "Wait a
4 second, that's off level." No. Actually if you
5 looked at the state test we did, the readability
6 analysis on the whole test and, for example, many
7 fifth-grade tests have readabilities that range
8 from mid-second grade to late fifth grade. Many
9 eighth-grade tests have readability ranges between
10 eighth grade and fifth grade. So, this is within
11 the range of the readability, so we did make it
12 easier to read. There's no doubt about it. And
13 the TAMI rating on this was 4.

14 Now, there are a lot of other aspects
15 behind this, and I'll illustrate some of them, far
16 more psychometric aspects in terms of actually
17 using this item that we went through.

18 So, if you do modifications, what do you
19 learn? This is a data table out of the Journal of
20 Exceptional Children. I think everything I'm
21 going to present today is in the published
22 literature, so you can access it and dig into it

1 deeper. There is one validity study that's under
2 review that I'll highlight.

3 But to orient you again, we looked at
4 students without disabilities, SWOD; students with
5 disabilities, SWD-NE, not eligible, so these are
6 kids with disabilities who didn't qualify for the
7 2 percent; and then these children who were
8 eligible, SWDE. If you process across this page,
9 and these are effect sizes, all kids got a sample
10 of items, equivalent form of items that were not
11 modified, and then they got a modified item. And
12 what we found is that the modifications improved
13 test performance of all students. The biggest
14 effect was for students who were eligible for the
15 2 percent assessment.

16 We also played with -- and because of
17 the accommodation issue about rating support, et
18 cetera, and these were all delivered online,
19 reading support does help, but the effect is
20 minimal for most students except for those who are
21 eligible.

22 Graphic visualization for those of you

1 who prefer to look at graphs. This is the graphic
2 for the testing condition for reading with the
3 children that are eligible but had the lowest
4 graph line. But yet you can see that the
5 trajectories of these three groups under the three
6 different conditions -- original, modified, and
7 modified reading support -- were virtually
8 identical with the greatest effect size for
9 students who were eligible for 2 percent.

10 Similar effects for mathematics, okay?
11 We were using some items here that actually came
12 from Discovery Education's online assessment
13 system but that they were considered to be
14 equivalent to our -- not psychometrically
15 equivalent but similar to items used in state
16 tests in Arizona and Pennsylvania and South
17 Carolina.

18 The evidence indices for desired
19 changes. So if I had the power to influence test
20 designers in various places, one of the things
21 that would be really great as we're going forward
22 is to have some common indices that we could share

1 out routinely and really document the science, if
2 you will, of the assessment process. So, these
3 are, I think, the desired outcomes. I've already
4 introduced these to you previously, this list on
5 the left, and in my way of thinking, here's the
6 evidence that we would look for.

7 So, if we want an increase in
8 accessibility, I would offer -- one approach would
9 be to use this tool we call TAMI, or something
10 equivalent, although I'm not aware of any
11 equivalent form right now. If we want to decrease
12 item difficulty, what do we look at? We'd look at
13 the P values that are 100 percent correct. If we
14 want to increase item discrimination, what do we
15 look at? We'd look at the point by Sheryl,
16 correlations.

17 And so the point is, I'm not going to
18 read all these to you, but for every desired
19 outcome, we actually do have the ability to
20 quantify or to characterize the effects and to
21 know what's happening. And I think we need to do
22 more of this and to share more of this so that we

1 can actually build a system that we understand to
2 change.

3 This is an important one that I want to
4 highlight here: Maintain DOK, depth of knowledge.
5 We used Norman Webbs' four-level system here, and
6 let's -- I want to highlight the fact that we can
7 make an item easier but yet have it assess the
8 same depth of knowledge that it was intended to.
9 So, we're not lowering it down to simply a recall
10 item to make it easier. We can stay within the
11 depth of knowledge. And we were able to do that
12 very well I think.

13 Number of accommodations. I think this
14 is serendipity that we found. The more we had
15 good item modifications, the fewer accommodations.
16 And from my experience with accommodations,
17 particularly in the lives of middle-school
18 students, we often find that even though they're
19 offered, they're actually not used, and the more
20 we can move to situations where the support that's
21 needed actually gets used, I think the better off
22 we are. That's the intention, that it helps all

1 of us along the way.

2 So, it would really facilitate, I think,
3 communication if we can start coming up with sort
4 of a taxonomy, if you will, of the indices of
5 change and the changes we're looking at. It's
6 tedious, but I know guys like Joe can do this. I
7 know it can happen. So, we're looking forward to
8 seeing more data out of this.

9 So, lessons learned from the item
10 modification research that we have. So, four of
11 them at least, and I've already mentioned the
12 first one, that there's persistence in this group
13 of reading difficulties. They're slow readers by
14 and large. And if you know anything about the
15 cognitive load literature, speed of reading
16 influences the memory process, too, and we had
17 kids -- God bless them -- reading for 15 minutes a
18 passage that might take 2 minutes for other kids,
19 and by the time they got to the end, they forgot
20 what was at the beginning. But yet they read
21 every word, okay? And so we just -- we have to
22 really keep those things in mind.

1 Modify items that are less complex. But
2 modified items are less complex but still aligned
3 with grade-level content standards and measuring
4 the same DOK can result, truly, in improved test
5 performance.

6 The data that I was showing you were
7 with fourth graders and eighth graders, too, as
8 well, so it's not just focused on one grade.

9 Easy and effective modifications
10 include, from my perspective -- again, I'm not
11 advocating any one of these as the package of
12 accommodations that we manipulated but had to do
13 with simplifying language, reducing the number of
14 answer choices, highlighting the questions asked.

15 I want to highlight that last point
16 about answer choices, make the answer choices
17 plausible, okay? If they're implausible, then all
18 that it becomes is more of a reading task. And we
19 learned when we looked at, actually, the answer
20 choices selected by children with and without
21 disabilities, that highly implausible ones were
22 rarely accepted, were responded to by children

1 without disabilities. Not so for kids with
2 disabilities. So, it becomes a little bit of a
3 trick, that we throw an implausible in. It
4 doesn't have any effect or has a negligible effect
5 on students without disabilities; it has a
6 disproportional effect on students with
7 disabilities. I know that's not a popular topic,
8 but that's a topic that we need to think about and
9 understand about the effect of and the
10 plausibility of the answer choices.

11 Modified items can result in equal or
12 better measurement precision, actually, and that
13 is a big-time goal for all of our work for the
14 eligible students, better measurement precision,
15 not just more accessibility but better precision.

16 So, conclusions from item modification
17 studies. Less is more. And there it is. I
18 really think less is more on that.

19 So, next area. So, it's actually the
20 easy work. The easy work is actually item and
21 test modification. The hard work has to do with
22 making sure that the inferences that we're making

1 from these test scores actually are valid. And
2 one of the key points has to do with instruction.

3 What if instruction on the content
4 covered by the test item has not been provided?
5 What if some students received instructions and
6 others have not received instruction on the
7 content tested? Well, a number of students and
8 also their teachers have told us repeatedly that
9 they had not been tested on the content on their
10 state test. So, we actually probed this. Part of
11 me says why the heck didn't I start here 15 years
12 ago? I started with testing and accommodation
13 research, but now I'm doing research on tests and
14 opportunity to learn.

15 I want to make one point in particular.
16 Test developers have usually tried to address
17 these concerns by conducting alignment studies,
18 sort of saying we've got the state content, we've
19 got the state tests. And they're aligned. The
20 contents are aligned. And of course they know
21 instruction goes on and that instructions can be
22 quite aligned or not so aligned. But we've used

1 some tools now to actually document the learning
2 opportunities. Here's a very blunt tool that we
3 used in Arizona: Simply asking teachers to
4 document whether they taught the content or not,
5 okay, and how much they emphasized it. This blunt
6 tool, though, turned into something a lot more
7 detailed. So, I want to highlight this.

8 What do we mean by "opportunity to
9 learn?" A lot of people use the term. It's in
10 federal legislation, access in general curriculum.
11 This is our definition: The degree to which a
12 teacher dedicates instructional time and content
13 coverage to the intended curriculum objectives,
14 emphasizing higher order cognitive processes,
15 evidence-based practices, and alternative grouping
16 formats. And just like with our work on item
17 accessibility, we found that we needed to invent a
18 tool in order to codify what was going on in
19 classrooms. So, we invented a tool called My
20 iLogs. Both these tools are available. They were
21 developed on federal grants. They're free to any
22 users that you want to use them.

1 So, this tool is called My iLogs. It
2 takes years of opportunity to learn research, and
3 it creates a unified model of opportunity to learn
4 where we look at time, the time that's involved in
5 instruction, the quality of that instruction, and
6 the content of that instruction. Teachers use the
7 tool, and they report to us what they do. We do a
8 third- party observation of these teachers along
9 the way.

10 Let me just pause for a second. If you
11 ask teachers to tell you what they teach, you
12 might think, well, they're going to -- if there's
13 going to be bias involved, they're going to
14 overstate it, okay? Probably not understate it.
15 So, think about that.

16 Here's the tool. It's a calendar-based
17 tool. All the content standards are on the
18 left-hand side. It's now set up for common core
19 standards. It's a drag-and-drop tool. Teachers
20 load the content they teach into the calendar.
21 They estimate the time they're going to use to
22 teach it, and they later go back and confirm that

1 time along the way. They tell you about the
2 content they taught, the cognitive level from
3 remember to create. They talk about the teacher
4 actions they take and the grouping processes.
5 Detailed cases available.

6 Here's an observation record that the
7 third- party observer goes in and watches the
8 teacher and provides a degree of agreement between
9 what teachers reported and what was observed.

10 Data that teachers get content covered across the
11 year. By the way, teachers use this daily in the
12 states of Arizona, Pennsylvania, and South
13 Carolina. We had teachers reporting over a 155
14 days of instruction on this, on both their class
15 level and individual students, and it takes 22
16 minutes a week to do it on average. That's a
17 week, not a day.

18 So, you get detailed data. And I'm
19 getting the time clock. But you can see the
20 content covered in a number of minutes that they
21 cover in each class. You can see some missing
22 areas. Content was not covered. So, you get all

1 sorts of feedback from this.

2 And in our work we always selected a
3 couple of students with disabilities to get
4 detailed feedback on for the teacher, and what did
5 we learn? Here's what we learned. Key findings
6 regarding opportunity to learn:

7 "Initial opportunity to learn data from
8 three states indicated that general and special
9 education teachers spent about 66 percent of their
10 allocated class time teaching the academic content
11 standards of the general curriculum; another 25
12 percent on custom skills; and about 5 percent on
13 non-instructional tasks and activities. Teachers
14 covered approximately 68 percent of the academic
15 standards based on the average of 151 school
16 days." We found no teacher in any of these states
17 that got better than 81 percent of the content
18 standards covered. We used a very low bar. One
19 minute counted for coverage. If you bump it up to
20 10, 20, 30 minutes, you actually see that
21 coverage, obviously, go down. This is challenging
22 work on the part of educators to cover the content

1 that we've asked them to cover.

2 "Teachers placed greater emphasis on

3 higher- order thinking skills in the general

4 education classes than in the special ed classes."

5 Not a surprise. Differences between general and

6 special ed teachers related to time on standards

7 and content covered indicated effect sizes of

8 about.5. Given that students in both types of

9 classes were held to the same general curriculum

10 standards, irrespective of the educational

11 settings, is problematic, the teachers in special

12 ed classes provided less instructional time and

13 less instructional coverage.

14 Major conclusion -- I know I'm out of

15 time, Susan -- but based on these samples, general

16 classrooms, which represented full inclusion model

17 -- kids with disabilities sitting in the regular

18 classes that Alexa talked about, about 60 percent

19 of our students getting the majority of their

20 instruction in these classes -- students with

21 disabilities experienced less time on standards,

22 more non- instructional time, and less content

1 coverage compared to their class.

2 At least for students with disabilities
3 nested in general education classes, opportunity
4 to learn appears to be a differential opportunity
5 structure. Structural differences do not indicate
6 equal or equitable opportunity to learn for
7 students with disabilities. Given this, we've --
8 given their disability-related characteristics,
9 students with disabilities may need at least as
10 much opportunity to learn if not more than their
11 peers without disabilities.

12 However, the current research findings
13 across these three states in middle schools
14 suggest that the exact opposite is occurring, and
15 if replicated these data would pose serious
16 instructional challenges for teachers and hold
17 profound implications for policymakers, focusing
18 on academic proficiency and growth without
19 consideration for the instructional inputs and
20 processes that affect these students.

21 This is published, this full report is
22 published.

1 Implications -- can I just leave the
2 slide up, Susan?

3 So, implications I think -- and maybe
4 this will start the dialogue amongst the people at
5 the table -- studies must be conducted to
6 establish the reasonableness of the validity
7 claims about the relationship between instruction
8 and test scores. We get that test score, we make
9 a big time inference that kids have been taught
10 this. And yet, in some cases, when they don't
11 know, it's because they haven't ever been taught
12 it, they haven't been exposed to it.

13 Teachers need substantial support to
14 meaningfully cover the intended general curriculum
15 with all students. Many students with
16 disabilities, if you take our data and project
17 out, would need 30 or 40 more days of class time
18 annually to have equitable opportunity to learn.

19 Finally, alignment is important, but
20 current alignment measurements -- and we use the
21 SEC and this one, the surveys of the inactive
22 curriculum at the end of the year and correlate it

1 with our OTL measures -- are insensitive to
2 individual students. They're not designed to do
3 that. Thus, alignment is a poor indicator of
4 opportunity to learn for students with
5 disabilities.

6 And I will stop there. I respect your
7 time, appreciate your listening to this. Thank
8 you.

9 MS. WEIGERT: Let's take time for one or
10 two questions, and then we'll move on. We have a
11 little extra time at the end of the session, so.

12 MR. FERRARA: Thanks. Steve Ferrara
13 from Pearson.

14 Steve, thanks for that great
15 presentation. I believe your results regarding
16 the differential opportunity to learn. However,
17 for us to be sure that we could -- that we -- that
18 there's validity to those results, we really need
19 what I think is one of the ultimate validation
20 studies -- would be looking at the correlations
21 between opportunity to learn and test performance.
22 Do you have any plans to any validation work like

1 that?

2 MR. ELLIOTT: We have some of that data,
3 actually, Steve, as well, and we're doing more
4 right now. We're engaged in it. And so, you
5 know, in the study that's under review right now,
6 we had last year's achievement; last year's
7 achievement still is the best predictor of this
8 achievement. But if you start looking at
9 opportunity to learn, it accounts for some
10 additional variance in that whole process. I
11 think that's what you're asking.

12 MR. FERRARA: Yeah, so some additional
13 variance.

14 MR. ELLIOTT: Yeah, it adds about 8
15 percent of additional variance.

16 MR. FERRARA: Right. Okay, thanks.

17 MR. ELLIOTT: Yeah. I'd be glad to
18 share that with you, too.

19 MS. FEDORCHAK: This is Gaye Fedorchak,
20 New Hampshire.

21 Steve, I appreciated your presentation
22 as well and your attempt to codify what elements

1 there are that make, you know, that we have to
2 consider when we're doing item development -- I
3 think some of Sheryl's comments earlier about, you
4 know, when accessibility needs to be built in.
5 One of the things that I can't say strongly enough
6 is that assessment and instruction is about
7 communication fundamentally. Deeply, most deeply,
8 it's about communication.

9 The elements that you described are the
10 traditional elements of an item. But when we're
11 thinking about kids with various disabilities,
12 it's very clear, especially from our APIP work
13 we've been very involved with, a number of our
14 states have, that item elements need to be
15 conceptualized based on the sensory cognitive
16 information processing that needs to be done in
17 order to demonstrate the cognitive construct. And
18 so there are going to be item elements that
19 conceptualize -- we're now moving into the age of
20 accessibility by item, not -- you know,
21 traditionally it's been by content area or by
22 subtest. You know, you do a calculator on this

1 part of the math test or not. It's the whole math
2 test, you know, or the whole reading test. We're
3 talking about item by item by item by item now.
4 And so we've got a whole different language
5 structure that we need to build around the
6 communication elements, both receptive and
7 expressive elements of an item, and interaction
8 with the test, you know, presentation, interaction
9 response, navigation. We've got to develop a
10 whole new nomenclature, and I appreciate your
11 work. We've got a lot more work to do.

12 MR. WILLHOFT: Joe Willhoft. Steve,
13 have you -- both the -- well, all of the
14 assessment consortia are designing assessments to
15 be administered online, so to what extent has TAMI
16 taken a look at that new item environment,
17 particularly, you know, use of screen size, tool
18 availability on the screen, and so forth and so
19 on?

20 MR. ELLIOTT: Good question, Joe. First
21 off, all of our research that we did in terms of
22 actually test performance of kids was done online.

1 But I can tell you, we used a product -- a
2 Discovery Education's product, which I don't think
3 was screen sized well and, actually, scrolling was
4 necessary and it interfered a lot. And so screen
5 size is a big deal. TAMI, per se, doesn't have a
6 category on that. I think that's sort of the next
7 step to modify that. The value of one item per
8 screen is terrifically important, and the value of
9 having embedded visuals that are attractive and
10 easy to understand are important. Science items
11 are very challenging with the visuals because of a
12 lot of labeling, et cetera, and I think there's
13 some great work, actually, done by John Sweller
14 with items. He didn't intend it to be items, but
15 it's terrific work that could be helpful in this.
16 And his work, actually, is largely done in
17 multimedia environments. I don't know if you know
18 the book -- I'm not here to sell his book -- but
19 it's Efficiency in Learning. It's really -- and
20 it's referenced in my handout.

21 MS. LOVING-RYDER: Good afternoon. Oh,
22 I guess it's still morning, isn't it?

1 In 2007, Virginia was awarded a GSEG
2 grant to develop a modified achievement standard
3 test for grade eight reading and mathematics and
4 also to look at the characteristics of students
5 who might be appropriate for this assessment.
6 Virginia, at this point, had a fairly mature
7 online assessment program, so what we decided to
8 do in developing a modified achievement standard
9 test was to look at how we could take existing
10 online items and add supports and simplifications
11 to those items using the online assessment. We
12 did involve teachers in identifying the supports
13 we might provide to the items. We did a number of
14 focus groups and also a number of surveys asking
15 teachers to tell us the types of supports they
16 provided to students and instruction that we might
17 then be able to mirror in an online environment.

18 In the process of developing these
19 assessments, we also did several pilots where we
20 did think-alouds with students, and some of the
21 other speakers have pointed out how useful this
22 is. But particularly when you're looking at

1 adding supports, it's very helpful to have direct
2 students' input on how successful those are. We
3 also continued to have teachers review the items
4 as we added supports.

5 At the time that we got the GSEG grant,
6 all of Virginia's items were multiple choice, so
7 our initial work was in looking at multiple-choice
8 items and providing supports to those. Since that
9 time, we have expanded so that we're now using
10 some technology- enhanced items that are not
11 multiple choice, and in the items I'm going to
12 show you, you will see some use of that item type
13 to provide supports to students.

14 In general, the assessment is about
15 percent shorter than the regular assessment. But
16 we did ensure that we have the same content
17 coverage. And for the multiple-choice items, we
18 did eliminate one distracter.

19 Based on the work of the GSEG grant, we
20 did expand our modified achievement standard
21 assessment so that we now will be implementing
22 mathematics for grades three through eight and

1 Algebra I. This year we're going operational. We
2 are field testing reading grades three through
3 eight, and of course reading this year with the
4 intent to go live next year. The reason we have
5 delayed full implementation is that we are
6 implementing these assessments concurrent with the
7 new assessments that are being provided to our
8 general education students.

9 I'm going to spend most of the time I
10 have demonstrating items and talking about the
11 supports we have provided, and just to give you an
12 idea of what you will see, I'm going to first show
13 you the item that was provided to the general
14 education students and the regular scan that's a
15 learning assessment. And following that, I will
16 show you the supported item used in the VMAST.
17 That's our Modified Achievement Standard Test.

18 I do have video clips within those VMAST
19 items that will show you the functionality. I
20 cannot promise that I will always synch up my
21 comments with the video clips, so just please be
22 understanding that it may not always synch up.

1 Okay, so with that, let me show you the
2 first item. This is a grade four math item, and
3 it's probably difficult to see, but it's an item
4 where students are asked to order the fractions
5 from least to greatest. And then this is the
6 supported item. Here, this is an example where we
7 have used a technology enhanced item to provide
8 support. The number of fractions has been reduced
9 to three, and you will see that students now can
10 use the drag-and-drop functionality to order the
11 items. And if they want to change their minds,
12 they can start again. So, this is an example of
13 the technology-enhanced items being used to
14 support.

15 The next item is also grade four math.
16 In this particular one, students are asked to add
17 fractions with unlike denominators.

18 And here in the supported item we have
19 provided a vertical presentation of the item and
20 then provided a workspace where students can
21 determine the common denominator. There's also a
22 hint that says to type the numbers in the boxes to

1 find the common denominator. You'll see, as I go
2 through these, that the hint boxes are very
3 common. They're always provided in the same
4 location, the upper right-hand corner. There is
5 never a requirement to answer the question but
6 instead provide some suggestions to the student as
7 to how they might go about solving the problem.
8 You can see here that the student still will have
9 to find the common denominator and answer the
10 questions. So, the construct is maintained, but
11 we are providing support to the student.

12 The next one is a grade five math item.
13 This is the original SOL item, and here the
14 student is asked to determine what the figure
15 would look like if the triangle, section 3, is
16 removed.

17 In the supported item, we did change
18 this quite a bit. Again, it does maintain the
19 construct. And here the student is asked to
20 identify what shapes will be left if the figure is
21 cut along the dotted lines. In this particular
22 one, there's another hint box. You'll see that

1 there's an animation button that they can push
2 that will cause the shapes to separate. But the
3 student still needs to be able to identify what
4 shapes are left.

5 Now we're moving on to grade six. This
6 is an item where the student in the original item
7 was shown six points on a coordinate grid, and
8 they were asked which of the three points can be
9 connected to form a right angle. The supported
10 item is changed quite a bit. Here we're asking
11 the student which point best represents the
12 ordered pair 5-1. And just as clarification, the
13 standard that this item actually assesses asks the
14 student to know the coordinates of an ordered
15 pair. So, this item actually maintains the
16 construct quite nicely.

17 Here the student is told that he or she
18 can place the pointer tool on the answer options,
19 and then it will show the lines of interception so
20 the student can determine which point best
21 represents 5-2.

22 This is one of my favorite items. I'll

1 just tell you that ahead of time. This is a grade
2 seven item, and the original item asks the student
3 to identify what is the most likely type of
4 transformation that takes place from figure 1 to
5 figure 2.

6 In the supported item, here we just have
7 one figure, and you'll notice that here the
8 animation is actually in the stem. So, in this
9 particular example, the animation is not a hint.
10 The student needs to click on the animation in
11 order to answer the question. So, if you click on
12 the animation, you'll see that -- it will happen
13 in a minute, I promise -- if you click on the
14 animation, you can see that the figure actually
15 transforms, and the student can continue to click
16 the button to see the figure transform again and
17 again. Still, the student has to understand that
18 this is a translation. So, it does not change the
19 construct.

20 Okay, the next one is another grade
21 seven math item. In this particular one, the
22 original item asks a student to determine the

1 volume of a cylinder, and here the student is just
2 given words to describe what he or she is to do.
3 In the supported item, you'll notice that we have
4 reduced the language load and also provided a
5 figure that shows what the cylinder looks like.
6 Again, we have a hint box, and we are providing
7 the formula.

8 On the regular SOL items, students are
9 provided with the formula sheet, but they have to
10 be able to find the correct formula and use it to
11 solve the problem. Here we have provided the
12 formula within the item, and then the student can
13 drag and drop the correct values into the formula
14 to determine the volume.

15 The next one is a grade eight item, and
16 it's a line graph, and the student is supposed to
17 determine if the height of the water continues to
18 decrease as shown, what will be the closest to its
19 height at 6 p.m. In this particular supported
20 item, we didn't change the content a great deal.
21 You notice that we did include a hint box. Again,
22 if the student holds the pointer tool over the

1 answer choice, he or she can see the extension of
2 the graph and then make a decision about what the
3 correct answer is. Doesn't change the construct
4 but does provide the student support in
5 visualizing where the graph will go.

6 This is a three-dimensional solid item,
7 and the student in the original item was shown
8 three different faces and is asked to identify
9 which of the three-dimensional figures could be
10 constructed using those faces.

11 In the supported item, we really didn't
12 add a lot of support except to add color. But we
13 did find that this was very helpful to a number of
14 students in helping them to identify where the
15 faces were and to put them together to construct
16 the figure.

17 Okay, we're moving on to Algebra I now,
18 and as you can imagine, thinking about supports
19 that we could use in Algebra I that would be
20 meaningful but would not change the construct of
21 the item was very challenging. This was probably
22 our most challenging test to identify supports.

1 In this one, the student is supposed to
2 identify what the equation for the lines shown on
3 the grid might be. This is the original item.

4 In the supported item, you'll see that
5 we have provided a name for the line, and then
6 there's a hint box where the student clicks the
7 animation button. It shows them the slope of the
8 line. So, the student can use that slope to
9 identify the slope, but the student still has to
10 find the intercepts to determine the equation of
11 the line.

12 And then moving on to another algebra
13 item, this is another item where the student is
14 asked to find the slope in the original item,
15 given two points. In the supported item, the
16 student is provided with a hint as to what the
17 slope formula is and then is provided with
18 information about how to plot points on the grid.
19 If the student plots points, they will actually be
20 connected so that the student can see the line,
21 again in an effort to assist the student in
22 determining what the slope would be.

1 Okay, before I move on to reading, this
2 is just a bit of a summary of what we found to be
3 helpful in terms of supports for math. The hint
4 boxes probably have been the most useful supports.
5 As you saw in the examples, they sometimes include
6 formulas. Sometimes they include strategies or
7 reminders. We've also found that it's very
8 helpful to color code important information. An
9 example that I didn't show you might be color
10 coding the variables and then the numbers that
11 would be substituted for those variables -- and
12 then the example of the grade eight item where we
13 colored the faces of the three-dimensional figure.
14 That proved to be helpful to students in
15 identifying the figure.

16 Online manipulatives, such as the
17 animation showing the translation. As we went
18 through these, you noticed that we did simplify
19 the numbers where we could without changing the
20 construct, simplified the language quite a bit,
21 and tried to provide graphics as support where
22 possible. And then of course in all of these we

1 did the multiple choice items. We reduced the
2 answer options from four to three.

3 Now, we are further ahead with
4 mathematics, because we have already field tested
5 and we're going live this year. We are still
6 working on determining supports for reading. We
7 will be doing, as I said, a large-scale field test
8 this year for reading. But, here, just a few
9 examples of what we've done in this area.

10 This is the original reading passage.
11 This is a grade eight one that would be given to
12 the regular students.

13 This is an example of the VMAS passage.
14 You'll notice that the text box above the story
15 describes what a flashback is. This particular
16 passage does include a flashback. It's not
17 something that's tested within the questions, but
18 it's something that would be helpful for the
19 students to understand. You'll also notice that
20 we have added a graphic with some color in the
21 hopes that that would be more engaging to students
22 -- and then some navigation tools so that the

1 student can page through the passage.

2 All right, the first item, the original
3 item, asked what the word "bantering" means in
4 paragraph 2. So, here is the supported item.
5 You'll see that what we've done is we've taken the
6 paragraph where bantering appears and presented it
7 with the item so that the student doesn't have to
8 go back within the passage and search for the
9 item.

10 The other thing we've done is that if
11 the student hovers on the particular option, it
12 will replace within the passage so that the
13 student there could use the strategy of word
14 substitution to determine the correct answer.

15 The next question -- this is the
16 original one -- asks the student what the italics
17 in the story means, and here is the supported
18 item. You'll see that part of the passage where
19 the italics appears automatically appears with the
20 item. And then the student can also use the
21 drag-and-drop functionality to put the answer
22 options within the stem to read through them and

1 see which one is correct.

2 In the last example I have, we didn't
3 provide you with original passage, but this is the
4 passage that the student would see on the VMAS,
5 and this is the supported passage. And you'll see
6 that this particular item deals with the passage
7 as a whole. So, in this case, there is a hint box
8 provided to remind the students that they need to
9 go back and look at the passage as a whole in
10 order to answer the question. Instead of having
11 to scroll back to the passage, they can actually
12 click on an exhibit window and the passage will
13 appear so that they can go back and review the
14 item.

15 And then, finally, the student, again,
16 can use the pointer tool to have the answer
17 options -- I'm sorry, actually, they can drag it
18 into the sequence here.

19 Okay, in terms of supports that have
20 been useful for reading, we shortened the reading
21 passages, but, again, as Steve mentioned, we were
22 careful to not reduce so much language that we

1 were changing the reading level of the passage.
2 Where an item asked about a particular section of
3 a passage, we did excerpts of the passage and
4 presented it with the item and provided hint boxes
5 when the student needed to return to review the
6 entire passage to answer the item.

7 We also had graphic organizers. It
8 simplified the language and not only highlighted
9 the words and the passage but also in the stem and
10 then in the case of multiple choice items reduced
11 the answer options from four to three.

12 We're also finding that the use of the
13 new technology-enhanced items that we're now
14 moving toward in Virginia has been very helpful in
15 providing supports for this population.

16 In terms of lessons learned, it is very
17 important to involve educators in identifying the
18 supports and also to, whenever possible, interact
19 with students so that they can tell you what works
20 and what doesn't. One thing we found early on was
21 that you have to be judicious in providing the
22 supports. It's very easy to over support, and

1 that does confuse students.

2 We have found that providing practice
3 with these items is imperative. We have practice
4 items for each of these tests on our website, and
5 the examples I showed you today are all examples
6 of those that are posted on our website. That's
7 the URL for the practice items there, and then
8 this is what the web page looks like. In Virginia
9 we also provide audio as an accommodation, so
10 you'll notice that we also have an audio version
11 of the practice items that students can use and
12 then a practice guide to go with each of these.

13 We have found that students need to be
14 guided through the functionality that is there, so
15 the practice guide is actually to be used by the
16 teachers to walk the students through the practice
17 items. We don't want to have a situation where
18 the student is just going through the practice
19 items independently, because then they may miss
20 the use of the hint box or the use of some of the
21 functionality.

22 I think application to the next

1 generation assessments -- of course, using
2 universal design is very important, reducing the
3 language load, and simplifying presentation where
4 possible. And then I think we need to maximize
5 the use of technology. There is so much that we
6 can now do with computer-based assessments to be
7 able to mirror the supports that teachers
8 routinely provide these students in the classroom.

9 So, thank you.

10 MS. BOWMAN: Thanks, Shelley, that was
11 great to see all the different item types and
12 where you guys have moved since you developed your
13 modified assessments.

14 I think one of the questions I know that
15 PARCC within our Technology Workgroup, our
16 Assessibility and Accommodations Workgroup, that
17 we've talked about in terms of looking at
18 students' various response modes and how they may
19 respond to different item types. Can you share
20 with us any insight you guys have learned in terms
21 of different students' response modes? There was
22 a lot of navigation, so how do we get at students

1 who may have those fine motor difficulties or may
2 need a scribe or someone else to help guide them?
3 Or, you know, in certain cases it may violate the
4 construct of a scribe in doing certain things.
5 So, could you speak to students with their various
6 response modes and how you entrust that with your
7 online assessment?

8 MS. LOVING-RYDER: In terms of the
9 navigation, that's where the practice becomes so
10 important. As we worked with students in the
11 think- aloud, we found that functionality needs to
12 always either be in the same place or have the
13 same kind of icon. For example, the animation is
14 always that little arrow. We still would provide
15 for a scribe if a student needs one. We do have
16 students who would have fine motor difficulty and
17 would have difficulty manipulating the
18 functionality. So, we would still allow for that.
19 Currently, we don't have supports that would be
20 based on an individual student's response mode,
21 but I will have to say that this is probably one
22 of the few assessments we have ever implemented in

1 Virginia that students actually like. They are
2 very engaged in it, and we were very surprised
3 that they liked it so much, so.

4 MR. HOCK: So, this comes from a
5 measurement person. I'm Michael Hock from
6 Vermont.

7 These are very cool. And the thing that
8 occurred to me watching this, though, is that in
9 many ways you made your test accessible by
10 providing what are just better items. And I
11 didn't see any reason why this is -- there's no
12 reason why our tests shouldn't look like this for
13 everybody, and I just wonder if you'd comment on
14 that. Is there any thought about Virginia's test
15 just being what your developing?

16 MS. LOVING-RYDER: Well, I think we
17 certainly are continuing to work with universal
18 design. You know, some of the supports we provide
19 probably are beyond what we would provide to a
20 regular education student. For example, we would
21 expect a regular education student to be able to
22 use the formula sheet, to find the appropriate

1 formula, and use it, but I do think we have
2 learned a lot in this endeavor about how to make
3 our items more accessible to students in general.

4 MS. FEDORCHAK: This is Gaye from New
5 Hampshire.

6 I just wanted to add to your comment.
7 If you were to explore this with kids in the
8 general assessment, one of the things that it
9 would allow you to do would be to see how kids in
10 the general -- you could collect data behind the
11 scenes and see how kids who explore the general
12 assessment use these tools, select part of the
13 formula that say, no, I don't think that's right,
14 and you can watch their error patterns and how
15 they explore the space and learn a lot about how
16 they're trying to approach the problem solution.
17 There's a lot to be learned.

18 MR. FERRARA: Does anybody know if
19 there's a point around this? All right, okay.

20 Well, good morning, everyone. Sorry to
21 those of you to my right whom I can't see and who
22 can't see me.

1 Susan, I want to thank you for giving me
2 the opportunity -- Susan and your colleagues, I
3 want to thank you for giving me the opportunity to
4 have some good, geeky, fun here today.

5 And also I have some -- you know, I live
6 here in the city, so naturally I was the last
7 person to show up and I apologize for making you
8 nervous. I'm from one of the new R&D centers at
9 Pearson. We're currently called the Center for
10 Next Generation Learning Assessment. I work with
11 a team on performance assessments with technology
12 applications, paper applications, and so forth for
13 assessing students, adults, professionals, and so
14 forth. And I'm looking forward to hearing the
15 kinds of comments that I'm going to get from my
16 psychometrics colleagues here at the table and --
17 I should say colleagues and friends -- and people
18 in the audience as well.

19 So, let me start by giving a -- talking
20 about -- an overview of what I'm going to talk
21 about. I'm going to make a brief comment about my
22 background and context for my comments; talk about

1 typical grade- level, live-scale assessment
2 design; say a few things about programs, one about
3 designing and developing modified assessments for
4 low-performing students; and suggest some ideas
5 where I think I'll go out on a limb -- and I'll
6 probably have a saw with me -- and then after I
7 propose those ideas, talk about what we need for
8 those ideas to work if they were acceptable to
9 people.

10 So, part of the reason why I love
11 talking about assessing students with disabilities
12 is I was a high school special ed teacher in
13 Massachusetts before I got into the measurement
14 field, the live-scale measurement field. I've
15 also worked on alternate assessments around the
16 country in several states and was awarded a GSEG
17 grant several years ago to work on modified
18 assessment -- actually designed research,
19 development, and tryouts.

20 So, let me start with -- I'm going to
21 take some liberties here and talk in very general
22 terms about how grade-level assessments typically

1 are designed and developed and what we typically
2 get from them.

3 So, here is -- this is real data from a
4 grade-level state assessment, and what you see
5 here is a 40-item test and the locations of the
6 items on the theta scale. Now, if I had a pointer
7 -- I'm afraid to mess with this. I don't want to
8 mess up the slides, but let me take one shot at
9 this. If I had a pointer I'd point out a couple
10 of details on this slide. First of all, at the
11 far left on the X-axis -- wow, look at that, like
12 magic -- you can see that the range of the scale
13 goes from negative 6 to positive 6. That's in
14 theta -- in the theta scale. I didn't want to use
15 the scale scores, because it might give away which
16 state assessment program this is. So, you can --
17 now, this is scaled using IRT, which means you can
18 portray items and examine these on that underlying
19 scale that, in this case, ranges from negative 6
20 to positive 6.

21 And you notice the item counts on the
22 locations on the scale. You can see, for example,

1 right there on the left is one item in the range
2 negative -- well, yeah, negative 2.9 to 2.0. You
3 can see that there are 15 items in the range
4 negative 9 to 0, and so forth. And, as I said,
5 this is a real test, so you can see there are lots
6 of items toward the middle of the scale and fewer
7 items on the tails. This is a very typical
8 design.

9 Now, what do you get out of this? This
10 is the same portrayal of those 40 items, and the
11 curve -- the dotted-line curve -- represents test
12 information. The inverse curve at the bottom is
13 the standard errors.

14 Now, the test information function or
15 curve gives you an indication of where the most
16 test information is. In this case, test
17 information is psychometric information that gives
18 you some assurance that you're getting a reliable
19 estimate of a person's location on the scale.
20 It's a form of -- it's a way of thinking about
21 test reliability. Its inverse is the standard
22 error curve, and that's those boxes in the low

1 dotted lines. I have a low dotted line at the
2 bottom. The advantage of the standard error curve
3 is that it's portrayed on that same theta scale so
4 that you can see in the middle of the scale where
5 those two bars of 15 items are, the standard
6 errors are quite low, and that's where the
7 information is the highest. And those standard
8 errors are actually around .2 on that theta scale
9 that goes from negative 6 to positive 6. So, the
10 standard errors are quite low.

11 When you go out to the tails where there
12 are fewer items, standard errors, the standard
13 error curve goes up, the standard errors get
14 higher. Again, this is a specific, real test.
15 It's also very typical. You see this typically in
16 state assessment programs. So, the standard
17 errors are higher. The score estimates are less
18 reliable than the parts of the scale where the
19 standard errors are lower.

20 Now, what does that mean in terms of
21 assessing students and saying something about
22 where they are on the scale?

1 Well, first of all, where the
2 information is highest -- again, where the curve
3 is high and the standard errors are low where all
4 the items are -- you can see that the targeting
5 here is right at and below the cut score, the
6 proficient cut score, and that's important because
7 that is the fundamentally most important
8 classification decision that has to be made;
9 that's where the consequences are highest; and
10 then the information is also high below that cut
11 score, because those are kids that are getting
12 close and you want to have a lot of information
13 there.

14 The difficulty for us with this typical
15 design for a state assessment program is when
16 we're talking about the so-called 2 percent kids,
17 and those are the kids who typically -- they are
18 at the bottom of the scale. They're defined that
19 way. They're at the bottom of the scale where the
20 information is lowest, the errors around -- the
21 confidence intervals around their scale scores, or
22 in this case their theta scales, are the widest.

1 That's where the fewest items are. So, at the
2 risk of sounding politically inappropriate, the
3 point is that grade-level assessments aren't
4 appropriately designed for lowest- performing
5 students; they're designed to get the best
6 information about the largest groups of students
7 and especially around the cut score -- the
8 important cut scores.

9 Now, having said that, what do we see --
10 to summarize? We see lots of precision for
11 students in the middle of the scale, less
12 precision for students at the upper and lower ends
13 of the scale. For this test, the overall score
14 reliability, alpha is probably .85 or higher. I
15 actually didn't -- forgot to get access to the
16 technical report, so I can't say for sure.
17 Typically, it's going to be at least .85. But that
18 overall reliability -- so, the result is good test
19 targeting for most examinees, less so for students
20 in the tails, especially our favorite kids here
21 today.

22 Whether it's been stated in measurement

1 terms or not, a goal of modified assessments was
2 to target tests specifically for low-performing
3 students with disabilities. And how do people --
4 what did programs do about that? They did things
5 like try to make tests easier. Several of the
6 other speakers have talked about that. And the
7 goal is to make items -- more items easier so we
8 can build test forms that are not just easier but
9 are appropriately targeted for students at this
10 low end -- currently at this low end of the test
11 scale -- the proficiency scale, at the same time
12 targeting on-grade content standards.

13 There have been several -- there are
14 several successful programs that have been
15 successful in doing that very thing, meeting this
16 psychometric and policy goal for assessments of
17 modified achievement standards. We saw a good
18 example of that just a moment ago. And there's
19 been a ton of research in the last five years
20 since the GSEG grants came out. Much of it has
21 been very productive. We saw a very good example
22 of that in Steve's presentation.

1 So, now we're moving into the next
2 generation, and the question is what are we going
3 to do about these kids we're talking about coming
4 back onto the grade-level assessments? And that
5 was the problem that I was asked to address.

6 So, what I'd like to do is propose a
7 couple of solution ideas. I don't think these are
8 the solution. The idea is to generate -- to
9 propose some ideas that people can shoot at and
10 start a discussion. And I'm going to talk about
11 two adaptive testing approaches. One is actually
12 implemented in several states around the country.
13 That's computer-adaptive testing. I should -- I
14 am -- I'm going to point out that that is
15 item-level computer adaptive testing.

16 Another solution that I'm going to
17 propose is another form of adaptive testing that
18 I'll call here today multi-stage, multi-level
19 testing. And I'm going to add a little twist to
20 it.

21 So, as I said, computer-adaptive
22 testing, the way it's practiced in many states --

1 and other assessment programs -- is item-level
2 adaptive test targeting. And what makes it both
3 adaptive and appropriately targeted for any
4 student who takes a CAT is -- you know how the
5 process works. Students respond to items, the
6 computer algorithm begins to estimate their
7 location on the scale -- again, that scale could
8 be the theta that I portrayed or the scale --
9 score scale that's used for reporting for program.
10 When a student is answering items correctly, the
11 difficulty of the subsequent items is increased,
12 because the indication is that the student is
13 higher on the scale than where the current
14 estimate is. And so students are given more
15 difficult items until they start to get -- they
16 start to respond with about a 50/50 success rate
17 on the items. And that's where the estimates --
18 the algorithm determines is the final estimate for
19 students. Likewise, if a student is responding
20 incorrectly, the items are gradually decreased
21 until the algorithm converges on the best estimate
22 for that student's proficiency.

1 There are quite a few benefits in
2 computer- adaptive testing that go along with the
3 many challenges involved in implementing
4 computer-adaptive tests for large-scale programs.
5 They include things like effective test targeting,
6 which was the point of my talking about this,
7 because in that process, the student sees many
8 fewer items that are above or below their
9 proficiency level than they would see on a
10 fixed-form test. So, it's efficient and
11 effective, and what it means is for, again, our
12 favorite students today they see many fewer items
13 than they see on a fixed-form test where they're
14 seeing lots of items that are well beyond what
15 they currently can do.

16 There are some interesting findings from
17 computer-adaptive and self-adaptive testing. I'll
18 say it very briefly. Higher-performing students
19 often say when they've taken a CAT or a
20 self-adaptive test that it's one of the hardest
21 tests that they've ever taken, and that's because
22 they don't get to see the easy items that are

1 aimed at the other kids. Likewise, the
2 lower-performing students say, well, that was one
3 of the best tests I've ever had the opportunity to
4 take, because they don't see the items that
5 they're not able to process and respond to.

6 All right, another approach that I'm
7 proposing for discussion here is multi-stage,
8 multi-level testing; and it's a form of -- think
9 of it as subtest or short test form adaptive
10 testing. The ideas here have been proposed in
11 many assessment design and measurement theory
12 books over the years. There are some operational
13 implementations. I think the reason why it didn't
14 get a lot of traction and show up in state
15 assessment programs is around the time when people
16 were proposing this, for example in Fred Luid's
17 1980 book. It was soon after that that computing
18 made it possible to do item-level adaptive
19 testing. So, this became less important to do.

20 One good example of an operational
21 multi-stage, multi-level test was used in the
22 high school longitudinal study of 2009 HSLs. The

1 cognitive test there in this longitudinal study
2 sponsored by NCES focuses on algebraic reasoning
3 and problem solving. Students take a 15-item
4 router test, and then depending on how they do on
5 that test, they're routed either up to a high
6 level, moderate level, or easy level of difficulty
7 stage 2 test.

8 Here's a sort of schematic illustration
9 that I borrowed a dissertation from 7 years ago,
10 and again if I had a pointer I'd walk you through
11 this, but notice there are three stages in this
12 schematic design. This is just one of many
13 multi-stage designs you could use. In this
14 illustration, all of the students involved or all
15 of the examinees take a test of average
16 difficulty, and it's going to be relatively short.
17 It's not going to be a full-length test. It could
18 be 15 items, for example, as in the case of HSLs.
19 Then, depending on how they do, they're routed in
20 stage 2 either to an easier test, a harder test,
21 or a test of similar difficulty.

22 And then there's another opportunity --

1 and through stage 2, you get another updated
2 estimate of how well they're doing. And so in
3 stage 3, they're routed again, either at the same
4 level, to a harder test, or to an easier test.

5 Now, how could we do this for a
6 large-scale state assessment of on-grade content
7 standards? This is where I think I might be going
8 out on a limb a little bit but, again, I'd like to
9 propose this idea for discussion.

10 So, this would be an example of a multi-
11 stage, multi-level, next-generation assessment
12 that would give us an opportunity to target test
13 difficulty for an on-grade assessment of on-grade
14 content standards for a state assessment program.
15 So, you start with an initial block of items.
16 Let's call that the stage one test, as I've
17 labeled in the display. And what we get out of
18 that is an initial estimate of the student's
19 performance. In stage two, examinees are routed
20 to a high-difficulty block, a moderate- difficulty
21 block or a lower-difficulty block.

22 Now, there are some details I need to go

1 over here. First of all, I'm suggesting that the
2 initial block -- sorry, all items in all four of
3 these blocks would be from a single grade-level
4 scale. So, all of these items in all of the
5 blocks are on the same scale. All of the items in
6 blocks are taken from a single grade-level pool of
7 items, which means you can compare scores from
8 each one of the blocks.

9 Second of all, that initial block would
10 be a representative sample of the full test
11 blueprint, and the reason why I think that's
12 important in general and in this case is because
13 I'd like to get an estimate of how low performing
14 students with disabilities are doing on the full
15 array of content standards that are covered in the
16 grade-level test. From there, you then have some
17 -- so, when you go to stage two and you route
18 kids, you now have some options. One is you could
19 simply use the stage two test to increase the
20 precision, enhance the precision, of the estimate
21 of their score that you got from the first block.
22 And you certainly can do that, you could do that

1 for the kids who go into the high and moderate
2 difficulty block. You could do that for the kids
3 who go into the lower-difficulty block if you
4 wanted.

5 I'm suggesting here that another way of
6 thinking about this would be to use that second
7 stage for students who you already know are either
8 through the stage one test or what you know about
9 their history. You could use that
10 lower-difficulty block in the second stage to do
11 some form of diagnostic testing.

12 An example might be -- so, here's how it
13 might go in a Race to the Top accountability
14 system. All the students take that initial block.
15 There actually could be multiple initial blocks,
16 if you wanted, at different levels of difficulty.
17 But using that initial block, you cover the test
18 blueprint, you get enough information to get a
19 reasonably reliable -- some people would say
20 unreliable -- estimate of people's scores. You
21 could use that for accountable reporting. Later
22 on I'll tell you why I think that might be a

1 reasonable thing to do. Then go into stage two
2 with our target students and, for example, either
3 based on what they did in block one or in prior
4 information that you have about where the
5 difficulties lie for these students use a targeted
6 diagnostic assessment in the second block. And
7 what that means is you could potentially get some
8 diagnostic information about a particular content
9 strand, for example, without increasing testing
10 time or the numbers of items that you'd have to
11 administer to students, and you'd still meet your
12 accountability requirements.

13 I'll pause here for a second and make a
14 point. I think I make it on the next slide, but
15 I'll foreshadow that anyway. The reason why I'd
16 like all of us to consider this notion is when we
17 give a fixed- form, grade-level assessment that's
18 targeted for the entire scale, all of the kids in
19 all the levels of the scale, we're already getting
20 unreliable estimates of how students at the low
21 end of the scale are doing. Likewise for kids at
22 the upper end of the scale. And so if you use a

1 shorter test in the initial block, you're likewise
2 getting an unreliable estimate, but it may not be
3 -- this is an empirical question -- it may not be
4 much worse than what you were getting in the first
5 place. And I'll show you a reminder of that a
6 couple of slides from now.

7 Okay, so having said that, what do we
8 need in order to be able to use item-level
9 computer adaptive testing or multi-stage,
10 multi-level testing in the ways that I've
11 described? What you need is lots of items, and
12 it's not just lots of any items. It's lots of
13 items that are on grade level; they're well
14 aligned to on-grade content standards. And for
15 targeting appropriately for the students we're
16 talking about, you needs lots of them covering all
17 the content standards located at the lower end of
18 that grade-level scale. That's a hard thing to
19 do.

20 When you look at many typical state item
21 pools for grade-level assessments, what you find
22 is there aren't a whole lot of -- unless the test

1 is already difficult for the state as a whole,
2 when you have a challenging test with challenging
3 performance standards, there tends to be a
4 sparsity of easy items and particularly for
5 specific content standards, which means it's
6 difficult to either operate a computer- adaptive
7 testing program or create these multi-stage,
8 multi-level forms.

9 There's another complication there that
10 I won't address in detail, but you also have to
11 deal with the issue of DOK. I think Steve Elliott
12 addressed that very nicely or made an assertion
13 about it.

14 So, how can we go about generating lots
15 of good items, enough items for appropriate
16 targeting? And here's another place where I feel
17 like I was going out on a limb. I'm not
18 addressing directly the advantages we can get from
19 technology enhancements. I knew that Shelley
20 would be covering that. And so I'm focusing on --
21 think of a practical situation where you're
22 working with your vendor or I'm working with my

1 colleagues and we have to generate lots of items
2 to do this appropriate targeting. We need lots of
3 high- quality items that are located at the low
4 end of the scale so we can do some appropriate
5 test targeting.

6 So, some ways to go. And I want to make
7 a point. I'm not taking existing items and
8 reducing their complexity in the ways that have
9 been talked about. I'm talking about developing
10 items for this specific purpose, new sets of
11 items. Well, you need some guidance on how to
12 develop items that are going to be located at the
13 lower end of the scale. One way to do that is to
14 look at the grade-level items that are located at
15 the lower end of the scale and analyze what
16 standards they're aligned to, what features about
17 those items we might consider easiness features.

18 Now, we know a lot from the research
19 that's been reported here today about those sorts
20 of easiness features that we can build all that
21 in. We can also use existing items as guides for
22 developing more items like that. We can also

1 develop cousins of those items, meaning if you
2 look at a set of items that are -- looking at a
3 particular standard, you can look at the
4 prerequisites of that standard and create a cousin
5 of the existing item as long as that prerequisite
6 still keeps the content on grade level. You can
7 do other things like reduce the complexity of
8 moderately difficult items by reducing cognitive
9 load and building in these other easiness features
10 that I alluded to.

11 Finally, you can also look at the kinds
12 of -- the complexity concepts in the Common Core
13 State Standards. When you go to Appendix A of the
14 Common Core State Standards, they talk about the
15 dimensions of reading complexity, and they talk
16 about -- they mention ranges within categories.
17 So, there are lower and higher levels of meaning
18 and knowledge demands, simple and complex
19 structure, and so forth. It seems reasonable to
20 think that in targeting appropriateness for
21 students with disabilities -- low-performing
22 students with disabilities -- you could focus on

1 that range of reading complexity.

2 So, let me make some closing comments
3 and, as I said, I foreshadowed some of these.

4 Well, this is one of the things I
5 foreshadowed. While grade-level test scores are
6 reliable overall, we get these grade-level tests
7 with reliability estimates of .85 and higher. The
8 score reliability for lower-performing students is
9 not as good as the overall test reliability
10 suggests, and that's because there are fewer items
11 at the lower end of the scale typically -- not
12 always but typically -- and fewer items at the
13 high end of the scale for the reasons I alluded to
14 before. And if you want to see evidence of that,
15 and you have a number of correct scoring -- your
16 program is a number-correct scoring system, you
17 can go to the scoring tables and you can see that
18 the standard errors for the scale scores at the
19 low end of the scale are very large compared to
20 the standard errors in the middle of the scale.

21 A second point: Multistage, multi-level
22 testing can work with fixed text forms. You don't

1 need to meet the challenges of item-level,
2 computer- adaptive testing, including computer
3 delivery. You can use fixed test forms and
4 deliver those test forms either online or in paper
5 and pencil. So, for now, where readiness for
6 online testing is still a challenge -- and I
7 suggest that's all over the place -- you can
8 continue with fixed forms indefinitely and provide
9 them either online or on paper and get the
10 advantages of multi-stage, multi-level testing.

11 One advantage of multi-stage,
12 multi-level testing compared to CAT is the
13 requirements for item pool size and replenishment
14 are not quite as demanding as for item-level CAT.

15 Now these ideas I've proposed for multi-
16 stage, multi-level design and item development,
17 there's a lot of work to do here. As I said,
18 these are some initial ideas that I'm hoping
19 people would be interested in discussing. And I
20 was asked to amplify what I mean by this, so
21 that's what I'd like to do in my final comments.

22 What do I mean by saying that there's

1 further conceptualization design and vetting
2 that's required here for this notion of a
3 multi-stage, multi-level testing? Well, one is,
4 is the policy community going to be willing to
5 accept an idea like this? It's not something
6 that's widely practiced in educational testing.
7 Where it is practiced, it's not well known. And
8 there's bound to be reluctance and resistance and
9 so forth to this idea, which is pretty well
10 developed in the literature, in textbooks, and so
11 forth but not widely practiced. And, you know,
12 there's been resistance and certainly challenges
13 that many of us have faced in implementing CAT, an
14 online testing for large-scale assessment
15 programs. So, there's a question as to whether
16 the policy community and state assessment programs
17 would be willing to stick their necks out or, to
18 use another metaphor, come out on the limb with me
19 with that saw that I have with me.

20 The other the thing is I also mentioned
21 that we need further research and experimentation
22 here, and that's related to this notion of

1 developing good quality items that we can use for
2 targeting appropriately low-performing students
3 with disabilities, students who currently are
4 located at the low end of state scales.

5 If I had an opportunity to do that sort
6 of work, I wouldn't go all in from the start.
7 What I'd want to do is lots of small tryouts, pick
8 ideas for developing targeted items, try them out
9 with kids, put them in a cognitive lab that has
10 been -- as Sheryl has recommended, try them out
11 with kids, look at how they function
12 psychometrically -- you can do that with classical
13 statistics; you don't need 1,000 or 1,500 kids to
14 do that -- and begin to build on the work that's
15 already been done and has been reported here and
16 build other ways of building these very large item
17 banks that are targeted for the students we're
18 talking about.

19 Finally, I'd say we already know a lot
20 about how to target items appropriately for
21 low-performing students with disabilities. What
22 we don't know a whole lot yet about is training

1 item writers to hit those difficulty targets. In
2 assessment programs where people have tried to do
3 that, the training so far has not been
4 particularly effective. I actually reported that
5 in a study that came out at the end of last year.
6 But there are very good ideas out there that have
7 empirical support in other applications so that we
8 could train professional -- not professional,
9 items writers -- we could train items writers,
10 whether they're professional item writers or
11 teachers in your states, to more accurately hit
12 difficulty targets, which enables us then to build
13 these large pools of good quality, on-grade-level
14 items that we can use for targeting at the low end
15 of the scale.

16 So, those are my ideas. Thank you.

17 MS. WEIGERT: Okay, can we have
18 questions for Steve?

19 Rachel, go ahead.

20 MS. QUENEMOEN: Steve might be surprised
21 to see how many of us are out on that limb with
22 him, probably sighing. But I think I'm going to

1 pick up on the very last statements he made,
2 because I'm not sure -- probably the most
3 difficult work we're doing in NCSC -- and maybe
4 it's more extreme for our group, but I don't think
5 so, I think it's just more obvious that we have to
6 do it -- is grappling with what makes for items
7 that truly get at the content that the kids are
8 learning in the way that they interact with that
9 content on the grade level.

10 Evidence-centered design is being -- the
11 words are being used in all four of our projects.
12 We certainly, at NCSC, have been on a steep
13 learning curve on understanding how that looks
14 differently from the way we've interacted with
15 content development before item development. But
16 everything hinges on us building a better
17 understanding of what that development within a
18 grade for all the kids in the grade looks like,
19 because if we get -- if it's only about difficulty
20 in a traditional test sense, we may have missed
21 it. So, what's occurring to me is, number one,
22 investing in understanding the content of grade

1 level and how kids get there -- and our friends at
2 DLM are developing multiple paths; we're
3 approaching it slightly differently -- and doing
4 that well is the absolute bottom line of whether
5 or not we're going to be doing better testing for
6 all the kids in the population. And it takes
7 time.

8 I absolutely agree with also getting out
9 there in small-scale tryouts in cognitive labs and
10 learning from teachers. Having a strong evidence-
11 centered design process to start with is just part
12 of the issue, because we don't quite understand, I
13 don't think, yet in the field what it looks like
14 to truly test across the broad distribution. And
15 part of that is that we've relied on our
16 traditional test development processes to inform
17 that.

18 So, number one is we're doing an awful
19 lot of upfront work in evidence-centered design.
20 Number two is how that interacts with our
21 traditional understanding of test design processes
22 I think is a brand new thing for most of us, too.

1 If we continue to develop tests the way we always
2 have and then just try to add in some work-around
3 content, we probably won't get to the yellow brick
4 road for all of us.

5 So, Steve, I'm just commenting that I
6 think there are lot of people noodling around out
7 there with you. The difficulty is that we have
8 old models and new understandings that are just
9 emerging, and we're on a tremendously fast
10 developing timeline.

11 So, any solutions? How do you fix the
12 rest of the system?

13 MR. FERRARA: Yeah, okay, so I didn't
14 address this explicitly or at least in any detail
15 in my slides. I alluded to DOK as a framework for
16 getting at the cognitive demands of items, and I
17 don't have any big problems with the DOK. It's
18 enabled people to think a little more explicitly
19 about the content knowledge and cognitive and
20 linguistic demands of items. Having said that,
21 there are more explicit frameworks under
22 development and available for use that get at, I

1 think, the kinds of stuff you're talking about,
2 Rachel. So, not only just focusing on making
3 items easier for kids but focusing on, again, the
4 content and cognitive and linguistic demands of
5 items. And so using other frameworks for focusing
6 on that sort of thing would be helpful, both for
7 training item writers and for evaluating items and
8 selecting items for the kind of test writing I'm
9 talking about.

10 So, this is shameless self-promotion.
11 If you look at the study that I cited here in the
12 slides, there's an example of another set of
13 frameworks that you can use for focusing on
14 content and cognitive, and linguistic demands and
15 becoming more explicit about those other things
16 that I think you're talking about, Rachel.

17 I'm afraid of what you're asking me.

18 MR. SHEINKER: You should be.

19 (Laughter) Alan Scheinker.

20 In thinking somewhat I won't say
21 traditionally, I appreciate your thinking out of
22 the box, and we were having a similar discussion

1 almost very similar to what you're talking about
2 today the other day, a couple of us. As we think
3 about in conventional blueprints that you create
4 and look at the coverage that the assessment might
5 have, how do you ensure in your model that you're
6 covering some breadth of coverage for our intended
7 population here and that we don't wind up
8 sacrificing time for opportunity, opportunity in
9 that they may not be exposed, even in the
10 assessment to the breadth of the standards? It's
11 the chicken or the egg to me that deals with if we
12 test, and we continue to test without changing
13 instruction, we'll still wind up with dipsticking
14 always. We know what they can do, but we haven't
15 changed anything. But that's --

16 MR. FERRARA: Okay, well, that was a
17 tough question. I'm sorry that I turned the mic
18 on, Alan.

19 (Laughter) So, you know, there's no
20 doubt that

21 Eventually when the Common Core
22 Standards are well mapped out in conceptual maps

1 or learning progressions and all that's validated,
2 that's going to be helpful in covering the range
3 of complexity at different levels of difficulty on
4 a test scale. I think that's a long time coming,
5 which is why I've talked about these item -- I've
6 alluded to these item-coded frameworks.

7 So, take a simple -- I think I'm getting
8 at your question -- take a simple example like
9 what I call answer-explain items. They're very
10 typical of short constructed response items where
11 you ask examinees to answer a question and then
12 explain it. They're applicable in all the content
13 areas. It turns out that item writers often write
14 a range of -- ask students to make a range of
15 different kinds of explanations, but they don't do
16 it particularly consciously or in a targeted way.
17 So, if you think of answer-explain items where the
18 explanation may require simply referencing the
19 specific word or sentence in a passage that
20 supports their answer or something akin to that or
21 other explanations where you have to defend your
22 answer or make a persuasive argument, I think the

1 complexity changes depending on the kinds of
2 explanations you require of students.

3 So, just using that as one simple
4 example, I think you could more consciously design
5 or develop answer-explain items that could be
6 where -- that could be located at the low end of
7 the scale -- I'm oversimplifying things here quite
8 a bit -- and give you an opportunity to cover a
9 range of complexity even though you're focusing at
10 the low end of the scale. So, okay.

11 MS. POSNY: I don't know about the rest
12 of you, but this has been probably, in my mind,
13 just the beginning. This has been very
14 intriguing, to say the least. And for those of
15 you who don't know, I have been involved in this
16 assessment process, especially for people with
17 disabilities, for probably at least 15 years or
18 more, probably starting back in the 1980s, so this
19 is something that -- you know, it's heartwarming
20 to know that we continue to work on it and
21 continue to do and try to do the right things for
22 kids. That is the bottom line. And we want to

1 continue to have the highest expectations for all
2 of our kids but realistic as well. And I think
3 that's kind of what we're grappling with.

4 First, what I'd like to do is I'd like
5 to say thank you, thank you to all of you as
6 speakers, thank you to all of you who represent
7 the Race to the Top assessments, the GSEGs, and so
8 forth. The questions have been great.

9 The other thing that I want to mention
10 is not only has this been informative, but this is
11 the beginning of what I see as a continuous
12 partnership as these assessments are being
13 developed. One of the reasons we brought, you
14 know, all of you together is to tap into the
15 knowledge and expertise that's here. You know,
16 the GSEGs paid for a lot of the development of
17 what we've done. Use that. I mean, you can hear
18 all the different pieces. They are here to help,
19 and that's exactly why we want to do that.

20 The other thing is that I want to offer
21 -- and I want to thank the audience as well -- and
22 I want to offer the effect to include

1 representatives from the disability community as
2 you're developing these assessments. I know that
3 they would just like to be able to know what it is
4 that you're thinking and to say wow, I think
5 that's a good idea or, oh, my god what are you
6 thinking? And that's pretty common. But I think
7 that's a good way to take a look at it.

8 As we continue to develop what I refer
9 to as the state of the art in terms of
10 assessments, we need to make sure that they're
11 valid, they're informative, and they really give
12 us the information that we want. And I think that
13 that's what we're all striving to do, knowing that
14 it's within a time frame that, yes, I remember
15 this discussion early on when we were talking
16 about when these assessments could be developed.
17 And I'll be honest with you, we thought, oh, maybe
18 by 2012, 2013, and I know now we're wrestling with
19 -- I know, we're wrestling with, you know, 2015,
20 and I thought, yeah, that's kind of the state of
21 the art. But, you know, it just takes a while,
22 and I think people understand that.

1 We need to continue to learn. And I am
2 looking to all of you to help us learn so that we
3 make sure that we really and truly are inclusive
4 of all students. And I think that, to me, is the
5 key and that's what I've heard here. We've heard
6 a lot of different ideas. I'm looking forward to
7 the state of the art and the continuous
8 conversations.

9 Thanks to all of you. Thanks to those
10 of you who accessed via the web, and this has been
11 a great beginning. Thank you.

12 (Whereupon, at 12:05 p.m., the
13 PROCEEDINGS were adjourned.)

14 * * * * *

15

16

17

18

19

20

21

22

1 CERTIFICATE OF NOTARY PUBLIC

2 I, Carleton J. Anderson, III do hereby
3 certify that the forgoing electronic file when
4 originally transmitted was reduced to text at my
5 direction; that said transcript is a true record
6 of the proceedings therein referenced; that I am
7 neither counsel for, related to, nor employed by
8 any of the parties to the action in which these
9 proceedings were taken; and, furthermore, that I
10 am neither a relative or employee of any attorney
11 or counsel employed by the parties hereto, nor
12 financially or otherwise interested in the outcome
13 of this action. Carleton J. Anderson, III

14

15 (Signature and Seal on File)

16 Notary Public in and for the Commonwealth of
17 Virginia Commission No. 351998

18 Expires: November 30, 2012

19

20

21

22