

U.S. DEPARTMENT OF EDUCATION

RACE TO THE TOP ASSESSMENT PROGRAM
PUBLIC & EXPERT INPUT MEETING

9:00 a.m.
Tuesday,
December 1, 2009

Capitol Peak Ballroom
Grand Hyatt Denver
1750 Welton Street
Denver, Colorado

FACILITATOR: JOANNE WEISS

PANEL MEMBERS:

LLOYD BOND
EDWARD HAERTEL
ROBERT LINGUANTI
THELMA MELENDEZ
JIM PELLEGRINO
GARY PHILLIPS
LORRIE SHEPARD
MIKE SMITH
ANN WHALEN

A G E N D A

<u>SPEAKER</u>	<u>PAGE</u>
Welcome and Introduction.....	3
 <u>Expert Presenters:</u>	
Edward Haertel.....	19
Lorrie Shepard.....	49
Gary Phillips.....	77
Lloyd Bond.....	96
Robert Linqanti.....	99
Jim Pellegrino.....	161
 Roundtable Discussion.....	 188
 <u>Public Presenters:</u>	
Randy DeHoff.....	271
Lindy Crawford.....	272
Alan Burke.....	283
Duncan MacQuarrie.....	287
Jim Ysseldyke.....	292
Clifford Lazar.....	298
Dennis Natali.....	303
Susan Zelman.....	308
Stuart Kahl.....	314
Jodi Papini.....	320
Stella Gibbs.....	325
Matt Gianneschi.....	331
 Adjourn	

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

2 (9:00 a.m.)

3 MS. WEISS: Good morning,
4 everybody. Let me ask you to come on in and
5 take seats. And welcome you today, and thank
6 you for coming.

7 I don't know how many of you were
8 here with us at the Boston event like this
9 that we did. But if you were, you know that
10 we have definitely come up in the world here.

11 I guess our government dollar goes farther in
12 Denver than in Boston, is the big learning
13 from today's lovely setting.

14 First, let me ask whether there is
15 anybody in the audience who needs interpreter
16 services? As you can see, we have somebody in
17 the front of the room, and we just want to
18 make sure that if you need the services, that
19 you are sitting down front where you can see
20 them. Okay. Great.

21 Well, thank you. I am going to
22 start by just giving you a quick orientation
23 to our agenda and our plan for the day. And

1 then I will stop and turn it over to the
2 people that we really came to this event to
3 hear from.

4 My name is Joanne Weiss. I am the
5 Director of the Race to the Top Fund of the
6 Department of Education. And I want to spend
7 just a couple of minutes at the front end
8 framing for you, why we are holding these
9 events and what questions we are trying to
10 answer as a result of them.

11 First, I also want to start by
12 making sure you know which event you are
13 coming to. This is not talking about the big
14 Race to the Top \$4 billion competition. That,
15 as you know, I think about two weeks ago, we
16 released the final regulations for that
17 competition. And that is a grant about
18 rewarding and encouraging states to undertake
19 comprehensive statewide education reforms.

20 What we are here to talk about
21 today is the Race to the Top assessment
22 competition, which is a \$350 million carve-out
23 from the larger fund. It is designed to

1 support consortia of states who are
2 implementing common standards by helping to
3 fund the assessments that might measure those
4 standards.

5 We are envisioning this as a new
6 generation of assessments, as you will see.
7 And that is what we are here to talk about
8 today. This competition is likely to have as
9 applicants consortia of standards. Like the
10 Race to the Top competition, 50 percent of the
11 funding has to be passed through to
12 participating LEAs.

13 So you will hear us talk a little
14 bit with our experts about how we might make
15 the best use of those funds. And the time line
16 for this competition, just like with the Race
17 to the Top competition, all funds have to be
18 obligated by September of 2010, because this
19 is part of the Recovery Act dollars.

20 We expect to have the regulations
21 for this out some time in March, and have
22 applications due in June. So that is the big
23 picture of this competition.

1 The goals for the state, for this
2 assessment program are several. First, we
3 want to support states in delivering a system
4 of more effective and more instructionally
5 useful assessments than we perhaps have had in
6 the past. We want to make sure that those
7 produce accurate information about what
8 students know and can do in terms of
9 achievement standards, in terms of individual
10 student growth, in terms of whether those
11 students are on track to be college and career
12 ready by the time they complete high school.

13 We also want to make sure that
14 these tests reflect and support good
15 instructional practice, and you will hear more
16 about this tomorrow, that they include all
17 students, including English language learners
18 and students with disabilities.

19 Now I should say that we have had a
20 separate sort of short topic format for a
21 number of issues that we thought deserved
22 special attention and a deeper dive by experts
23 who had speciality in those areas. We had one

1 in Atlanta right before the holidays, on
2 students with disabilities. We have one here
3 tomorrow on English language learners.

4 But the big takeaway that we hope
5 everybody gets from this is, that although we
6 are doing deep dives into those couple of
7 topics, because we think that there is some
8 specific expertise we need to tap into, it is
9 a conversation that you will hear us talk
10 about a lot today as well. We are really
11 trying to think through how we design these
12 assessments in a way that makes them
13 accessible to all students, particularly
14 English language learners, and students with
15 disabilities at the front end, rather than
16 having it be an afterthought that is tacked on
17 at the back. So it is stuff that you will
18 hear us talk about today, even though we are
19 going into it in deeper detail tomorrow.

20 We of course need these assessments
21 to be able to inform a number of different
22 decisions. And one thing that I want to just
23 put on the table and acknowledge is, that we

1 are doing things in a somewhat strange order
2 of things, thanks to the opportunities in
3 funding that we have on the table.

4 We are designing assessments,
5 knowing that we have an accountability system
6 under No Child Left Behind today that we need
7 to be compliant with. But also knowing that
8 EFCA is up for reauthorization, and we don't
9 yet even have the proposals on the table for
10 what that will look like.

11 So we said here that while we are
12 not talking about accountability today, we are
13 talking about assessment, we do need to be
14 cognizant of the fact that the outputs of
15 these assessments, the data that comes out of
16 these assessments needs to be able to inform
17 teaching, learning, and program improvement
18 decisions, determinations about school
19 effectiveness, about teacher effectiveness,
20 about principal effectiveness, determinations
21 about individual student, college and career
22 readiness. So this last bit on this slide is
23 our nod to the fact that we need to be

1 cognizant of the purposes for which these
2 assessments will be used, even as we are
3 designing them without specific accountability
4 framework in mind for the future; only the one
5 that we have now.

6 Which also leads us to some of the
7 requirements that we have put in place for
8 this, that we need to comply with the current
9 framework, even as we go forward. So that at
10 a minimum, these assessments need to cover
11 reading, language arts and mathematics. We
12 need to cover grades three through eight and
13 high school.

14 We are seeing that there is
15 summative assessments that we found over the
16 last few weeks, as we have been having these
17 conversation that this word "summative" is
18 used in all different ways by all different
19 people. I knew "formative" was used that way,
20 but now I am finding that "summative" is as
21 well.

22 And so by summative we didn't mean
23 that it was given only at the end of the year.

1 It might be given in fact, at other times
2 during the year. It doesn't mean that it is
3 only given once during the year in this
4 concept. It could be given at multiple times
5 throughout the year. It also doesn't mean
6 there is necessarily only one test.

7 We know that we have a number of
8 different goals here, and we are talking about
9 how to design a system of assessments that
10 might make the most sense, in order to provide
11 the information we are interested in. So it
12 might consist of more than one test. And
13 these are things that our experts will be
14 speaking directly to, over the course of the
15 day.

16 The other thing that we have said
17 is, that these assessments can be thought of
18 as assessments that may replace rather than
19 add to the assessments that are currently in
20 use. And of course, they need to be valid,
21 reliable and fair. In the notice that we
22 published in the Federal Register announcing
23 these meetings, we actually have a more

1 extensive list of their requirements. But I
2 did want to just pull out those few to talk to
3 you about today.

4 A couple of other things that I
5 wanted to go into. Why are we having these
6 meetings? Those of you who know how these
7 regulations work, the process that we are
8 going through here is not our typical process
9 for how we decide new regulations that come
10 out. Typically, they would be developed at
11 the Department of Education.

12 We would put them out for a formal
13 30-day public comment period. We would take
14 in those comments, figure out what we think we
15 should do, what we thought we should do as a
16 result of it, make changes to it, and put out
17 the final regulations.

18 As we sat to put pen to paper for
19 this competition, to design what those
20 regulations might look like, it became really
21 clear that we needed to have a different kind
22 of input at the front end, from experts. This
23 wasn't necessarily a question that was, what

1 does the public think. This is a question
2 where we have a lot of expertise in the
3 country, that we really needed to tap at the
4 front end as we were thinking through and
5 designing the competition.

6 We also felt like there were a lot
7 of different points of view out there. And we
8 needed to have a conversation that helped us
9 understand whether there was some convergence
10 on what new generations of assessments should
11 look like.

12 The other thing that all those of
13 you who have read regulations know all too
14 well is that they are a terrible device for
15 actually communicating a vision. They come
16 out looking more like a laundry list than a
17 picture of what could be.

18 So one of the goals we also had for
19 these meetings was that they would help us
20 paint a vision for what assessment, what new
21 generation of assessments that actually
22 provided useful and structural informations,
23 and that the parents looked forward to

1 receiving in the mail, and that teachers used
2 to make instructional decisions. What that
3 generation of assessments could and should
4 look like at this sort of 30,000-foot view.

5 And we have also asked our experts
6 to then dive down to the ground and help paint
7 a very concrete picture in response to the
8 notice that basically says, if you were us,
9 what would you do? Answer the question very
10 specifically as well, so that we are trying to
11 both paint a vision and get very specific and
12 concrete input from experts around the country
13 about how they think assessment systems should
14 be designed to meet some of the goals that we
15 have put out there.

16 The other thing that we know very
17 very well, and it is why many of you are
18 sitting up front at the tables here, is that
19 the applicants for this competition are the
20 states. And the states are suffering from
21 exactly the same sort of noisy information and
22 dearth of information all at the same time,
23 that we in the Department are feeling.

1 And so we wanted to make sure that
2 however we organized this, it was not only for
3 the benefit of the Department, it was also for
4 the benefit of the states. And that all of
5 you were able to hear and learn along with us.

6 We have also put -- let me come
7 back to that in a minute. We have also put at
8 your -- when you signed up, when you signed
9 in, I think you got index cards. And we urge
10 you, as we are going through the conversation
11 to write any questions that you have on index
12 cards, and you can drop them back at the front
13 table.

14 There is also a bunch of people in
15 the room. Mark, if you want to just stand up.

16 You can just come and give the cards to,
17 bring them up here, and we will use that as a
18 way to get your questions asked and answered
19 throughout the course of the day as well.

20 Let me go back. And then I will
21 come back to some of the other quick
22 housekeeping things. But the agenda for the
23 day has expert presentations. They are going

1 to be going in half-hour blocks per
2 presentation. Where they have got 20 minutes
3 of presentation and then ten minutes of
4 clarifying questions that those of us at the
5 table up here will be asking. You are on your
6 own for lunch. And I think you have got
7 recommendations of different places that are
8 nearby that you could run out to for lunch.
9 We will start back up here right at 1:15
10 sharp. Have a few more, a couple more
11 presentations after lunch. Then we will have
12 our round table discussion, a quick break, and
13 then we have an hour for public input. People
14 signed up ahead of time. And we have twelve
15 public speakers who are going to be giving
16 quick presentations to us during the 3:45 to
17 4:45 spot. So that is what the agenda looks
18 like for the day. In addition to feeling free
19 please to submit your questions through the
20 day, as you have them, so that we can feed
21 them into the conversation. You may notice
22 that I will be a bit of a ruthless timekeeper,
23 because we just do feel it is important to

1 make sure that we are hearing from everybody
2 that we have asked to present today. We have
3 handy little timekeeping devices up here that
4 should help me from having to be too ruthless.

5 But we will do that. And when you are doing
6 the public speaking, you also see that you
7 have got a timekeeping device on the podium to
8 help you keep to the time limit that you have
9 got. Please do put your cell phones on
10 vibrate. And we want to also let everybody
11 know that the session today is being
12 transcribed. It will be posted on our website
13 within a couple of days. All of the
14 presentations that you are seeing here today
15 will also be posted on our website. All of
16 the written input that we have received so far
17 is already posted on our website, and we want
18 to remind you, that if you have additional
19 written input that you would like to submit
20 for our consideration, the deadline for that
21 is tomorrow. So run back to your computer and
22 hit send on that, if you haven't done so yet,
23 if you have something that you would like to

1 share with us.

2 I do want to take a minute to thank
3 the states that have traveled, in some cases
4 long distances, to be here with us today. And
5 I also want to assure you that Lorrie Shepard
6 from the great State of Denver has assured us
7 that despite the fact that we are going to
8 watch a beautiful snowstorm roll in today,
9 until there is 25 inches, nobody should worry
10 about getting out of the Denver airport.

11 And with that, what I would like to
12 do is just start quickly by going around the
13 room, and letting all of us up here introduce
14 ourselves quickly to you, and then we will
15 kick off the first presentation. So Ann, do
16 you want to?

17 MS. WHALEN: Ann Whalen with the
18 Office of the Secretary and the Department of
19 Education.

20 MR. SMITH: Mike Smith, in the
21 Office of the Secretary.

22 MS. MELENDEZ: Thelma Menendez de
23 Santana, Assistant Secretary of Elementary and

1 Secondary Education, Office of Elementary and
2 Secondary.

3 MS. WEISS: I am Joanne Weiss,
4 Director of the Race to the Top Fund.

5 MR. HAERTEL: Edward Haertel,
6 Stanford University.

7 MS. SHEPARD: Lorrie Shepard,
8 University of Colorado at Boulder.

9 MR. PHILLIPS: Gary Phillips, the
10 American Institute for Research.

11 MR. BOND: Lloyd Bond, recently
12 retired from the Carnegie Foundation.

13 MR. LINQUANTI: Robert Linquanti,
14 West Ed.

15 MR. PELLEGRINO: Jim Pellegrino,
16 the University of Illinois at Chicago.

17 MS. WEISS: So thank you to
18 everybody. I want to end with a quick
19 commercial for our website, where there really
20 is a lot of information already posted that we
21 have gathered over the course of the last few
22 sessions. And with that, I am going to send
23 this over to Ed and let him get going.

1 MR. HAERTEL: Thank you, Joanne.
2 It is an honor to be here. And I am pleased
3 to have a chance to make a few comments that
4 might be helpful as the Department proceeds.

5 I won't read this quote to you.
6 You will have a chance to read it in the
7 website, if you don't have time to do it right
8 now. But it talks about the need for higher
9 standards. About the fact that the
10 assessments that we have now tend to focus on
11 low level skills. And it enforces a pattern
12 of low expectations.

13 This quote may have a contemporary
14 ring. For some of you, it may be familiar.
15 It is in fact from the Report of the National
16 Council on Educational Standards and Testing
17 from 1992. So that raises the question at the
18 top of the slide; what is different this time?

19 There are some new ideas. We have
20 heard a lot about common standards, and that
21 is important. Standards that somehow or other
22 avoid the thicket of trivia that many of our
23 standards end up calling upon our teachers to

1 address, and try to focus on a smaller number
2 of key ideas, deep, important ideas.

3 There is talk about alignment.
4 Individual growth is mentioned there, because
5 we are developing infrastructure now to use
6 student level data linked across time at the
7 state level, for entire state systems, which
8 is impressive, and will be useful.

9 There is explicit attention to a
10 range of different assessment types and
11 formats. Content sampling trends and so
12 forth. And there is also a bullet there for
13 21st century technology which includes new
14 methods of collecting data, as with computers.

15 New statistical models, as with longitudinal
16 growth models. And new psychometric models,
17 as within an incentive design.

18 There are also some old ideas. And
19 rather than discussing these new ideas from
20 the previous slide, I would rather talk about
21 three old ideas that I am going to contend
22 must be rethought. These are all
23 interrelated. And you will see how they need

1 to be considered in concert.

2 The first old idea is that if
3 assessments are aligned to standards, then
4 growth and test scores will track the intended
5 schooling outcomes. Alignment is important.
6 We have talked about it for a long time. It
7 is kind of a mantra. We need to think a lot
8 harder about what it means, and what all needs
9 to be included under that umbrella of
10 alignment.

11 Second old idea is that this is
12 intended to capture our common places about
13 what testing looks like. What we are doing
14 when we take a test.

15 Testing means students are working
16 in isolation. They are working, using what
17 they have inside their heads without access to
18 resources. They are answering clear questions
19 with clear answers. These are simply the
20 things that go with our ideas of what taking a
21 test means.

22 And finally, the third old idea is
23 that we can build one assessment system that

1 is going to do it all. As Joanne said a few
2 minutes ago, we know that we need to have a
3 system that is going to meet a lot of
4 different needs.

5 I have listed just a few bullets
6 here from the various documents that were sent
7 out in advance. We have a system that is
8 going to track progress at different levels
9 and inform decisions for individual students.

10 Informing solutions of higher education about
11 the qualifications of prospective admits.
12 Improve teaching and learning in the
13 classroom. Determine the effectiveness of
14 schools and principals and so forth.

15 These various different purposes
16 bring with them different design constraints.

17 And that means that the multiple indicators
18 that are included in the system can't just be
19 multiple forms of items or multiple formats.
20 It has to be something more than that that is
21 multiple about it.

22 I will skip over this fairly
23 quickly. These theories of action are if

1 anything, less clear than they were in 1992.
2 We need to think clearly about what is
3 supposed to change as a consequence to the
4 assessment system, and connect the dots to
5 make sure there really is a clear and coherent
6 story, part of a sensible story, a supported
7 story that gets us from here to there.

8 This thought shows us the way in
9 which a lot of it is thought of now. And
10 envisioning for example, in NCLB. We start
11 with academic content standards, which every
12 state has adopted. Those content standards
13 are supposed to direct both curricula and
14 instruction and instruction and also the
15 assessments.

16 And the theory is, that if both C
17 and I and assessments are aligned with content
18 standards, then they will be aligned to each
19 other. This is more the way it works in
20 practice. The content standards are in fact
21 linked to the assessments.

22 We have processes for assuring
23 alignment formally. It is in the RFPs. It

1 states not for testing companies to build
2 their tests. And those assessments in turn
3 are what largely drive curricula and
4 instruction. But the problem with the
5 narrowing of curriculum and instruction is
6 just what is tested arises.

7 Because with this pattern of
8 linkages, the curriculum and instruction may
9 end up being a poor reflection of the content
10 standards. I picked just two items, and
11 associated standards from the state
12 assessment. These are from a public website.

13 Many states release items so that people can
14 see what tests are about.

15 It would be tempting to spend more
16 time than I have just critiquing the standard,
17 which is ambitious but has a little bit of a
18 laundry list feel to it. You see that we are
19 talking about discussing, note the verb, a
20 wide range of content and so forth.

21 I am going to show you an item that
22 relates to that, but tests that standard.
23 This is formally aligned. The state says on

1 the website, this is the standard this item
2 tests. As you see, in order to solve this
3 item you need to have, to know certain
4 important things like Alliance for Progress,
5 Kennedy; Fair Deal, Truman; Great Society,
6 Johnson; New Deal, Roosevelt. If you know
7 that, then you can answer this, and a whole
8 slough of other items that look a lot like it.

9 Another example, high school
10 biology. Students note the big ideas here,
11 storage of energy, new structures created,
12 energy dissipated as heat. An important sort
13 of set of ideas for the college. Here is a
14 corresponding item. The answer is shark. I
15 could substitute food chain for energy pyramid
16 here. And in this case, none of the really
17 deep ideas from the standard end up being
18 tested in the item.

19 Now these are not in fact randomly
20 chosen. But I didn't have to look very hard.

21 So in summary, it is not enough for each item
22 to be somehow connected to one or more
23 standards. We need to think about alignment,

1 not just of standards and assessments, but
2 also curriculum and instruction. More pieces
3 need to be brought into that alignment puzzle,
4 and considered in concert.

5 A couple of possible ways of doing
6 this will be to use curriculum-imbedded
7 formative assessments, or to build on the good
8 work that has been done over the last couple
9 of decades on using portfolios. They need to
10 real specified.

11 We need to put some constraints on
12 what is going on in the classroom, because
13 there are certain things we need to have, if
14 we are going to have any information about
15 students that is really useful for assessment.

16 But it can't all be the externally mandated,
17 on demand, dropped from the sky kinds of
18 assessments.

19 The next old idea that needs to be
20 rethought, about the common places of testing.

21 Now these have been -- there has been a lot
22 of rethinking of this old idea for a long
23 time. I have cited three references here.

1 Norm Frederickson in the "Real Test Bias,"
2 *American Psychologist*, 1984, talked about
3 well-structured versus ill-structured
4 problems.

5 The fact that on tests we have
6 problems where there is a well defined
7 solution space. There is a single right
8 answer. There is a solution process that is
9 guaranteed to give you that right answer. And
10 unambiguous criteria for telling what that
11 answer is.

12 Gale Baxter and Robert Glaser,
13 1998, talked about -- in an article on
14 investigating the power and the complexity of
15 science assessments, talked about a quadrant
16 of different kinds of assessments where this
17 quadrant, content leaning process constrained
18 refers to items where you don't need to know a
19 lot that is deep about the way the content is
20 linked together and built into a coherent
21 structure.

22 And the process that you step
23 through in doing the task is sort of

1 cookbooky, in a step by step. We are talking
2 about performance assessments here. These are
3 not all multiple choice tests. But even
4 performance assessments aren't just
5 automatically better. They need to be
6 carefully designed and thought through.

7 And finally the idea of sequestered
8 problem solving comes from John Bransford and
9 Dan Schwartz. There is an excellent chapter
10 on rethinking transfer. From the *Review of*
11 *Research and Education* in 1999, and by
12 sequestered problem solving, Bay and Ian, prop
13 tests where students are working in isolation
14 in a constrained environment.

15 We do this on purpose. The
16 psychometricians say, to standardize is to
17 increase comparability and objectivity and
18 reliability. And there is a place for those
19 kinds of assessments. They are not
20 automatically bad.

21 My point here is just they are not
22 enough. So rethinking what it means to take
23 the test would mean expanding our idea of

1 assessments to also include in addition to
2 those kinds of tests, which measure things
3 that are important, also including some ill
4 structured problems with multiple solutions.
5 Also including these content rich process open
6 kinds of tests, where it is not clear what the
7 solution path is, and there are multiple
8 places that students might end up, that are
9 better or worse along a variety of different
10 dimensions.

11 It also includes what Bransford and
12 Schwartz referred to as tests of preparation
13 for future learning. Seeing whether students
14 can use what they know to figure out how to
15 approach a new problem. Figure out what it is
16 that they need to get smart about in order to
17 do that, to solve that problem successfully,
18 and where they go to find that information.

19 This final, my third point is about
20 rethinking the idea that one test can do it
21 all. A single form of testing can't satisfy
22 the full range of Race to the Top
23 requirements. And there are distinctive

1 design constraints. I am sorry. Distinct
2 design, I need to back up. Let me go forward.

3 A comprehensive assessment system
4 has to include multiple forms of evidence. It
5 can't all be externally mandated tests, not
6 just end of course, but even with the interim
7 assessments, tests that are given on a fixed
8 occasion and involve the same kind of evidence
9 gathering aren't going to be enough.

10 We need to have students engage in
11 a range of performances that are of interest,
12 including finding, organizing, and evaluating
13 information. Working together sometimes.
14 Working on unfamiliar problems sometimes.
15 Presenting what they know through written
16 reports and other kinds of presentations.

17 Dylan William summarized the need
18 for this range of assessments, to quote him
19 crudely, we start off with the intention of
20 making the important measurable. And end up
21 making the immeasurable important. This is
22 the story of where the testing has been done.

23 And it is the reason why a lot of indicators

1 has to go well beyond just different kinds of
2 formats on our tests.

3 Different kinds of assessments,
4 different kinds of uses of tests, were
5 referred to in the guidance for this meeting.

6 Referred to in Joanne's presentation, have
7 different design constraints. The design
8 constraints in terms of turnaround time. The
9 forms of which information is provided.

10 And the granularity of the
11 assessments, the range of content that is
12 covered or sampled are very different for
13 tests to inform instructional decisions in the
14 classroom, versus tests that are used for
15 accountability or for tracking individuals
16 through the progress over an extended period
17 of time. If you want to know whether a
18 student is on track for career or college
19 readiness, that is a different kind of
20 question from the kinds of questions that
21 teachers need. And they call for different
22 kinds of information.

23 Different kinds of linkages -- I am

1 sorry. We need to link standards of
2 assessments both horizontally and vertically.

3 There is a lot to talk about vertical scales.

4 And the particular challenge of that, because
5 of the way the NCLB regs are. But that is
6 something we should be able to finesse.

7 We also need horizontal linkages.

8 We need some connections across the subject
9 areas that students work in. Reading is not
10 just taught in English Language Arts. Reading
11 is taught in the counting areas and so forth.

12 And we need to accept the fact that some
13 parts of the assessment system will not
14 provide individual student scores.

15 Matrix sampling is a powerful tool.

16 It is used in important applications,
17 including NAEP, TIMMS and other international
18 assessments. It has been used in some state
19 assessments. It was used in California in the
20 old California assessment program for about 14
21 years.

22 But it does not allow us to give
23 individual student scores, based on all of

1 these test questions that students answer.
2 And for that reason, it has been a hard sell
3 for these kinds of accountability
4 applications, but it is an idea that needs to
5 be brought back, because it allows us to build
6 down some of these kinds of assessments that
7 are inherently more costly and less reliable,
8 and just use those for data at the aggregate
9 level and not necessarily at the individual
10 level.

11 In summary, we have to rethink
12 alignment, not just standards, test to
13 standards, but also bringing in these other
14 components of the system. The alignment has
15 to include curriculum and instruction. It
16 needs to include the assessments.

17 And I might add, it needs to
18 include teacher preparation. One of the
19 reasons why we end up with the kinds of
20 standards that we do, and the reason why we --
21 our social pressure is towards standards that
22 are a mile wide and an inch deep. You know,
23 not a Mile High. It comes from the fact that

1 teachers rely on particular ways of thinking
2 about curriculum, rely on certain kinds of
3 curriculum materials.

4 And there are going to be immense
5 pressures, to, even if we have fewer clearer
6 standards that really go for big deep ideas,
7 there is going to be all kinds of pressure to
8 give teachers materials that are like the
9 materials they have now, that they can use to
10 teach, in ways that are like the ways that
11 they teach now. And for that reason, the
12 alignment really has to go beyond just getting
13 the standards and the tests and even
14 curriculum and instruction together.

15 We need to look at how we prepare
16 people to use these things as well.
17 Concerning what it means to take a test, we
18 need to go beyond just sequestered problem
19 solving. Just putting, you are sequestering
20 students. Putting them by themselves and
21 having them work, show us what has been stored
22 in their heads by popping it out again.

23 A more complex view goes to

1 different ideas about what knowledge is about.

2 It means giving students more complex tasks
3 that are judged by multiple criteria.
4 Sometimes, having to work things out over an
5 extended period of time. Sometimes, having
6 them work together. Sometimes requiring them
7 to figure out what information they are going
8 to need and where to get it.

9 Again, this is not to say that the
10 kinds of tests that we rely on most heavily
11 now are bad or wrong. They are just
12 insufficient. They leave off too much and you
13 know from experience, again and again, that
14 what they left, what is left out of the tests
15 ends up, too often getting left out of
16 student's education.

17 Finally, we must rethink the idea
18 of multiple indicators. This is not just a
19 broader mix of item formats. There are calls
20 at NCLB for multiple kinds of evidence, but
21 that has been construed and interpreted
22 through regulations, which we all love, as you
23 said, in ways that don't really force us to

1 bring in the kinds of performances of students
2 working together at things I was just
3 describing in the last slide.

4 In order to accomplish this, some
5 portion of classroom assessment need to be
6 sufficiently documented and structured so they
7 can be brought into a comprehensive assessment
8 system to meet these multiple needs, so the
9 system is going to have a lot of moving parts.

10 So thank you very much.

11 MS. WEISS: Questions?

12 Sure, go ahead.

13 MS. MELENDEZ: Ed, in what role do
14 you see teachers in the development of the
15 assessments, especially the curriculum-
16 imbedded assessments and the formative
17 assessments that you have talked about.

18 And you also mentioned teacher
19 preparation. But what about professional
20 development? And how can the alignment and
21 the rethinking of the alignment play a role?

22 MR. HAERTEL: There's one old idea,
23 and it is a good idea that working on

1 assessments is good in service preparation for
2 teachers. I can find that idea going back to
3 Tyler's-eight year study.

4 We can find it featured prominently
5 in the new standards project, the Resnick and
6 Resnick study in the early '90s, that when
7 teachers are forced to think through what it
8 is that kids are actually supposed to be able
9 to do as evidence that they learned what there
10 is to be learned, that that is a good process.

11 Bringing teachers into that
12 process, and also, having teachers score
13 student work; not -- sometimes not just work
14 from their own schools but the work of
15 students at other schools. It can help to
16 communicate expectations and standards and
17 give students, give teachers good ideas.

18 There are some other systems in
19 other countries where this is relied on to a
20 larger extent. And with the appropriate
21 social supports and induction of teachers and
22 to those ways of doing the work of teaching.
23 Those systems can work quite powerfully.

1 It is not enough to count on a new
2 kind of assessment to give us new -- to give
3 us professional development on the cheap. It
4 is not going to work all by itself. It can be
5 a useful tool, a useful resource.

6 But I think we need to look hard at
7 all of the places where teachers are getting
8 their messages about how they teach.
9 Unhappily, the most important and most
10 powerful messages are, every one of our own
11 experiences in school as children. We have
12 all been there, and experienced the \$15,000 or
13 whatever.

14 And we know what classrooms look
15 like. And it is hard to shake people loose
16 from that. But certainly, having teachers
17 participate in assessment development is one
18 part of that. A risk -- a couple of risks
19 that I have seen in having teachers
20 participate, one is -- and this comes out of
21 experience with the Golden State exams in
22 California, was that teams of teachers were
23 invested in working on particular assessments.

1 And there was not a lot turnover
2 over time, and a sort of in group, in a few
3 cases, developed a sense of ownership over the
4 assessment and became resistant to dropping
5 items, the psychometrician said weren't
6 working; for example, resistant to developing
7 a large enough item pool that we could have --
8 to meet the needs.

9 So teachers can be valuable
10 resources, but we can't just hand it over to
11 them. The other problem is that teachers tend
12 to judge the quality of instructional
13 activities by a particular quality, as
14 instructional activities. It is not always
15 the same thing as the value that has
16 assessments.

17 So teachers can be valuable members
18 of the team. But there are other skills that
19 are required to build the easements that they
20 don't have, and shouldn't be expected to have.

21 MR. SMITH: Ed, let me pursue on
22 alignment for a minute. Imagine that you are
23 going to construct a set of assessments. And

1 there isn't -- there are common standards.

2 What steps would you take? What do
3 you -- what is your thinking about how to
4 create that alignment with assessment, with
5 the curriculum. Does the curriculum have to
6 come first? Does the assessment have to come
7 first? Do they have to come simultaneously?
8 How would you do it, if you were a director of
9 this?

10 MR. HAERTEL: Yes. I think I would
11 start. I don't have a good answer. But I can
12 tell you where I would start looking for an
13 answer; Preparation for Future Learning. The
14 way we do it now, is we build a curriculum
15 framework, which really is a guide for
16 curriculum and instruction. A set of common
17 standards that are ambitious.

18 We tend to do this through a broad
19 consensus process. And the easiest way to get
20 consensus around the table is to give
21 everybody what they want. So these standards,
22 documents, tend to become very broad. And the
23 politicians can point to them and say, look

1 how broad and ambitious our standards are, and
2 everybody is happy.

3 Then we turn around with another
4 piece of legislation or other regulations and
5 say now you have to use this framework for
6 curriculum and instruction as a test
7 specification. And for test specifications,
8 we need something else. So we finesse it in
9 the alignment process.

10 That is the place where these two
11 pieces come together. And we have decision
12 rules for deciding whether each item can be
13 mapped to some particular standard. The two
14 examples I gave you, you can see how the
15 mapping can be done. We have review
16 committees after the fact that look at that,
17 and we end up with something that is less than
18 satisfactory.

19 The test items -- no one can point
20 to a particular item and say kids don't need
21 to know that. Every item seems to be testing
22 something that matters. When we take those
23 items together, as a whole we see what we have

1 ended up with is a pure reflection of what we
2 were hoping for.

3 So I guess what I was thinking
4 about alignment differently, I begin with the
5 standards. And I tried to substantiate a
6 couple of alternatives exemplars of learning
7 units that would show what those standards
8 meant. And then for those learning units,
9 they think about how, what forms of
10 assessments, what forms of evidence I could
11 rely upon to learn how students would --
12 whether students were really getting the
13 important components.

14 Then in some cases, that would be
15 an exclusively saying what has to be left out
16 of the antecedent instruction because there
17 has to be something new about testing
18 something that is genuinely important. So I
19 might deliberately reserve some applications
20 or some instances or something for the
21 assessment context. So I guess that would be
22 the order.

23 I would start with the standards.

1 Then I would look at how those standards would
2 look in an ambitious curriculum without the --
3 and then after that, I would look at what
4 forms of evidence I might be able to assemble
5 and what pieces I need to build into my
6 comprehensive assessment system to get there.

7 MR. SMITH: So these are
8 performance standards of some sort that you
9 have created? Examples of good student work?

10 MR. HAERTEL: They are the
11 benchmarks. Yes.

12 MS. WEISS: So Ed, let me just sort
13 of put you on the spot, and say, so based on
14 all of these guidelines, and rethinking that
15 you have been doing over the last years, but
16 even a few, a couple of weeks that you have
17 been thinking about this, what are -- so if
18 you were designing this system, how many tests
19 would there be?

20 How frequently would they be given?
21 What would you specifically do with all of
22 this?

23 MR. BOND: Don't ask me that

1 question.

2 MS. WEISS: I am making a note, ask
3 Lloyd --

4 MR. HAERTEL: If I was designing an
5 RFP, I would try to give states flexibility on
6 that. Probably the question is what kind of
7 constraints would you need. There have been
8 different models that have been proposed, and
9 the right model is going to depend on the
10 state's own history and context, the size
11 resources.

12 I think that they are probably is
13 going to be some kind of end of course
14 summative assessment. But I would hope that
15 there would also be assessments, either
16 portfolio-based assessments or learning units
17 that we have assessment components associated
18 with them that could be slotted in.

19 There might be some agreement of
20 teacher choice on that. Those might or might
21 not be tied to existing curriculum packages.
22 But we have -- we are developing psychometric
23 technologies now for being able to bring

1 different kinds of evidence from different
2 potential assessments into a system, and
3 update our priors on what each student knows.

4 As with evidence and a design, so
5 it is possible to build a system where
6 different assessments could be brought in and
7 calibrated to an existing set of parameters
8 and put under existing scales and use in a
9 flexible way. And you can actually get
10 teachers a fairly substantial menu of options.

11 That is a little bit blue sky. We
12 are not really quite there yet. And it would
13 be a mistake to insist that people use this
14 technology in a large scale before it is
15 really ready for them. That is the direction
16 I would hope that we would be heading.

17 I haven't given you a very clear
18 answer to your question. But I'm a professor,
19 so I get to do this.

20 MS. WHALEN: Can I ask a follow-up
21 question that is right between Mike and
22 Joanne's. Given that alignment is really
23 central to one of your premises and also given

1 that we are thinking about this as states
2 coming in as consortiums applying as a group,
3 what are some of the preconditions you would
4 want to see in an application about this
5 alignment? So what would you need to see this
6 come together, agree upon, have evidence of,
7 as part of the application?

8 MR. HAERTEL: I would hope for some
9 common agreement about curriculum and
10 instruction, and the ways in which what
11 constraints or -- yes, constraints teachers
12 are going to accept to bring their classroom
13 instruction into alignment and provide the
14 kind of evidence that is required to inform
15 the decisions of the system.

16 You need to have buy-in. You need
17 to have agreement to opt for at least a
18 demonstration site submitted by the states, or
19 all of the schools participating to provide
20 data in the form that the system required.
21 One nice thing about our current system is
22 that even though it is regarded as oppressive,
23 it actually makes pretty limited demands of

1 what is done in classrooms.

2 MS. WHALEN: You would change that.

3 MR. HAERTEL: I would change that a
4 little bit.

5 MS. WEISS: Great. Lorrie,
6 let's -- I am going to pass the device.

7 MS. SHEPARD: That is Ann's
8 question. Ann's question is a perfect segue,
9 because if you wanted to have a preorganizer
10 for what I am going to emphasize is, how
11 important curriculum is, and our RFP
12 characteristic would be that people have to
13 have thought in advance, in their application,
14 about how they are going to deal with
15 curriculum.

16 Because we won't make progress in
17 this new vision if we don't have a plan about
18 that. In talking today, I want to acknowledge
19 the National Academy of Education working
20 group on standards, assessments and
21 accountability. Here are their names and a
22 website where you can get this white paper
23 recently released.

1 I owe this entire group a lot of
2 the thinking in my presentation. And I will
3 try to signal when I am off on my own. You
4 won't see the red and blue masthead when I am
5 trying to claim their authority behind what I
6 am saying. I will use it.

7 I, in relation to Ann's question,
8 am arguing that the Department needs a theory
9 of action for how they expect assessment
10 reform to occur profoundly in this go-round.
11 Why it needs to be different from 20 years
12 ago, when we attempted many of the same
13 things.

14 And importantly, I not only think
15 we don't know enough to have a single test and
16 system, I am arguing here that I don't even
17 think that each consortium can take on its
18 entire across the grades, across the subject
19 areas, assessment system and do well, what we
20 are proposing to be done, in which case then,
21 the right size to get all the way to the
22 ground has to be envisioned, so that people
23 take on a manageable task, and they have to

1 understand the politics of engaging this as
2 well as the technical and logistical --
3 logistical means things like just how much
4 waste of teacher time is it to score beyond
5 the umptyumpth one, because there is -- of
6 course, I would argue later for a teacher
7 learning. But at some point, it is a burden.

8 It is not learning. So there is -- that is
9 what I mean by the kinds of deep concrete work
10 that needs to go on to get this right. And
11 you need your applicants to think, what could
12 we take on?

13 Now that doesn't mean you leave
14 everything else the old way. So I am
15 suggesting either through reauthorization of
16 the SEA, which is not on the table today, but
17 in that first bullet there, I am thinking
18 there has got to be some other way to think
19 about getting all the systems up to the best
20 we currently have.

21 And you might use national
22 assessment as an example of sort of the better
23 version of the current state that does resolve

1 some of the problems that Ed outlined. But is
2 it all the way to what we are talking about
3 here today. So this is an argument about how
4 to signal what we do know and leave room for
5 uncertainty.

6 So what Ed answered about
7 performance assessments is really one of the
8 things that I will argue for, I am arguing
9 against this particular vision that came up in
10 Boston. Don't imagine that you are going to
11 carve up the United States and get some small
12 number of consortia to sign up for this stuff.

13 I am thinking two neighbors are going to have
14 trouble.

15 You are not going to get this size
16 consortia to do the work that we are talking
17 about in answer to Ann's question. For here
18 is my one slide about changing the character
19 of the assessments. And I am just taking off
20 from everything that Ed has already said.

21 The problem with alignment is that
22 it allows a cheat. It allows people to take a
23 matrix usually, a processes and content

1 strands. And then if the items fit somewhere
2 within, they say it is aligned. And we either
3 need to declare and all put our hands in the
4 middle, and agree that alignment isn't going
5 to mean that anymore.

6 Or we have to have different word.

7 And I have used the word "embodiment" in an
8 evaluation of NAEP's Mathematics Assessment;
9 Feldero used the reach and range. But the
10 idea is, you have to have a fulsome
11 representation of what you want students to be
12 able to do. Otherwise, we have lots of
13 evidence about the distorting effects.

14 In terms of what you should require
15 in the RFP, versus what you should allow,
16 again, using Ed's words, the flexibility to
17 the states, I think it should be a certainty
18 given a certain state of the -- the current
19 state of the art. Bless you. That
20 performance assessments are part of what is
21 ventured upon and engaged.

22 But whether it should be only at
23 the end of the year, or scored to count for

1 accountability or not, I think that should be
2 tested out in different ways. I hope you
3 actually treat this as a natural experiment
4 and get some variety on those things for which
5 there is some uncertainty, and where we have a
6 great deal to learn from trying this out.

7 I think that in the previous
8 conversations in Boston and Atlanta, there is
9 pretty much agreement that people are talking
10 about a system of assessments. And they are
11 thinking about formative, benchmark.

12 I will raise some questions about
13 that and summative. And when I say get the
14 system right all the way to the ground, then
15 instead of doing math and reading and science,
16 only doing math in a segment and seeing how in
17 a particular application it should, the
18 summative assessment should relate to the
19 formative assessment is the completeness you
20 want to push for. That you do want people to
21 have thought through this.

22 They will do it in different ways,
23 because as I have said, there is not one right

1 way to get the formative assessments aligned,
2 really aligned with curriculum so that then
3 the test is fair. We are going to have a lot
4 of issues about fairness, because if you
5 haven't had a chance to learn it, it is not
6 fair and that is what brings us to curriculum.

7 You know, when you had generic low
8 level skills, you could say it was fair to
9 everyone. But if you want people to be able
10 to engage real content knowledge, use their
11 reasoning in the context of content knowledge,
12 critical thinking isn't content free, well,
13 you have to think about what are we going to
14 teach and provide so that everybody has a
15 chance to do it.

16 And how those are going to link
17 across the levels is part of what getting a
18 sample assessment system implementation is
19 what you are trying out with this notice. My
20 questions about benchmark is, yes, I can see
21 how we can do it. We can go into that further
22 in the questioning period if you like.

23 But remember, what we have now was

1 invented in response to NCLB. And it is the
2 worst example of current practice. So please
3 don't just keep calling for it. I think you
4 can't do without it now, because school boards
5 want to see them. They want to know sort of
6 in October how things are going.

7 But remember, right now this is
8 driving the high-stakes low-level skills focus
9 even worse than the research we had done
10 previously with just the summative tasks. So
11 there is a lot of -- I have a lot of cautions
12 around including that in what you presume in a
13 testing system.

14 The importance of curriculum is
15 clear, based on some of the arguments I just
16 made about fairness and giving kids a chance
17 to learn it. But it certainly is apparent if
18 you look at Bill Schmidt's and other
19 researchers' work internationally. This is
20 what other countries do that we want to be
21 like.

22 We do not have an organized way of
23 saying, this is what we are sure you are going

1 to master in fourth grade. And by the way, it
2 has to be a manageable set. So what that is
3 already said about the politics of everybody
4 tacking on their own thing, led to the mile-
5 wide and inch-deep curriculum, never agreed to
6 as a curriculum. It means that everybody is
7 teaching different things, driven by the
8 textbooks.

9 Taking different things from within
10 the textbooks, because of course, they are so
11 big, and weigh so much, that each teacher has
12 to choose their own way through. The common
13 core standards are one step toward addressing
14 this issue. But common standards are not
15 enough because they aren't at the level of
16 curricular detail.

17 There is a next level of what you
18 are actually engaging the students to master.

19 Like when do you want them to get the
20 understanding about the relationship between
21 fractions and ratios, for example. So
22 curricular, deep in the meaning of standards
23 and provide both students and teachers with

1 the way to mastery. This is what developing
2 competence looks like it, and you have a plan
3 for it.

4 And here, I am citing the work of
5 the National Research Council Committee on
6 Science arguing for the horizontal coherence.

7 That is what Ed showed you the picture of.
8 Linking standards, instruction and assessment
9 for vertical coherence that is the state
10 better have some relationship to what is going
11 on at the next lower levels of district,
12 school and classroom.

13 And for developmental coherence,
14 this leads to the part of my talk that is
15 going to be about learning progressions. How
16 do you actually get smarter in mathematics, in
17 each of the strands. What do you do? First,
18 ensure mastery, what do you move on toward?
19 And that is what other countries have that we
20 do not have.

21 I would like to point to the
22 connected mathematics project as one instance
23 when NSF invested in developing curriculum.

1 Interestingly, they separately invested in
2 assessment design, and had they done the two
3 together, Mike, I think they would be
4 answering your question.

5 Your question to Ed, I would say
6 you do them together you have, because what we
7 need to actually consider, my last sentence in
8 this slide should directly consider
9 generalization and transfer. If you think
10 about it, you could actually map out, here is
11 the first set of tasks. And they look a lot,
12 they would look very similar to the students.

13 But what do students in my class
14 know? As soon as they get this, I am going to
15 ask it another way. And they have to learn
16 that it is not going to be a pat, familiar
17 format, and what I want to have on the table
18 when I am designing is all of the ways that I
19 could ask them.

20 And in fact, now I am going to take
21 that knowledge and I am going to use it in an
22 unrelated application. And furthermore, on
23 the final exam at the end of the year, it is

1 fair to ask you this kind of question. That
2 range of generalization is part of curriculum
3 building if you think of the problem of
4 transfer as being very real.

5 And recognize, we were asked not to
6 talk today too much about sort of the sinister
7 criticisms of the current system. But you
8 have to have them in your mind. You have to
9 know what teaching the test looks like when it
10 leads to productive learning.

11 And what does teaching the test
12 look like when it doesn't lead. It leads to
13 nongeneralizable. The kids can't even do it
14 if you ask it a slightly different way.

15 If you appreciate that research
16 then it schools you in how to design the
17 opposite. How to design the more fulsome
18 representation for assessment purposes, and
19 even how you stretch instructional
20 expectations so that you are always pushing at
21 that, and not letting kids stand pat.

22 This is my political speech. And I
23 think one of your speakers at another session

1 said, agreeing on even scope and sequence
2 won't be easy. I think we have to worry about
3 what it is going to be like for consortia to
4 agree.

5 So I have offered some suggestion
6 here, that focusing on just one content area I
7 think will help you get the logistical and
8 technical issues clear. But it will also be
9 easier for states to say, okay, okay, you can
10 force us to agree on writing. Skill areas, I
11 predict are going to be easier than content
12 areas like mathematics.

13 Because you are going to have
14 people saying, what, we all have to buy the
15 same textbook? Not necessarily. But that is
16 what is going to be at risk here. That is
17 what the choices and the decisions that people
18 are going to have to make are going to make
19 them think I am not willing to give up on
20 that.

21 Especially in content areas, the
22 problem of agreeing on curriculum could be
23 softened by saying, Okay, let's just agree on

1 six big ideas and assess those in common,
2 because we are literally experimenting and
3 evaluating here what it means to work on
4 common curricula.

5 Several of my slides, but I only
6 have five minutes left. I argue about the
7 importance of learning continuum. And if you
8 would hold this up, you will see an example of
9 a leaning progression, from Australia. No,
10 hold it up. You have to open it up. Think
11 of --

12 (Simultaneous discussion.)

13 MR. BOND: Now we can see it.

14 MS. SHEPARD: The most important
15 thing about learning progressions that I could
16 say today, that isn't being said adequately in
17 the previous hearings is, you must understand
18 the difference between substantive learning
19 progressions and statistical trajectories.
20 And I love Colorado's growth model. It is
21 statistical trajectories.

22 And it tells you a lot about who is
23 keeping up and who is not. But it doesn't

1 tell you why. And learning progressions go
2 further. They do have a statistical
3 underpinning, that there has been more work
4 that goes into them to say what those essays
5 look like on that Australian example.

6 And by the way, that state, that
7 national or in that case, provincial
8 assessment has the same trajectory that is
9 used in the classroom. And the kids can see
10 that when they are scored a certain way, where
11 they are on that continuum. Think of it as a
12 criterion-referenced growth model.

13 It is substantively referenced, and
14 we are going to have to do more work to pick
15 and anchor the tasks substantively that also
16 satisfied the statistical criteria. But you
17 can't just walk out on the street today and
18 get that. And please don't think you are when
19 people sell it to you.

20 While much research is still
21 needed, it is possible to start with
22 judgmentally set sequences, and improve them
23 empirically. So we heard a lot of talk about

1 oh, we don't know enough. Spin the sciences
2 in particular and the NSF-funded projects.
3 They are doing micro-developments of learning
4 progressions.

5 And those aren't the same as scope
6 and sequences arches of expectation. And we
7 want to do both. And we even want to know how
8 they relate to each other. But you don't have
9 to wait until the last NSF project is
10 concluded to think about learning progressions
11 in your assessment design.

12 This would be an example of one of
13 the benefits of a substantively grounded
14 learning progression for teaching. You can
15 actually know in a unit of instruction what
16 some of the classical misconceptions are. And
17 this is one, when kids are learning about
18 magnets, and they have to learn about what
19 makes the -- what things are going on inside.

20 And they almost always, when you
21 ask them to draw a model of this magnetized
22 mail, I think that all of the positive guys
23 are at one end, and all the negative guys are

1 at the other end. And the classic question,
2 the challenge instructional question is to ask
3 the kids what happens when you cut the nail in
4 half. And immediately, they know those red
5 guys couldn't run fast enough to get at the
6 positive end of the cut. They have to revise
7 their model.

8 And this kind of instructional
9 resource is what you get when you build
10 curriculum and assessments concurrently. I
11 would also like to argue that we are not
12 giving enough attention in this work,
13 especially because we keep asking
14 psychometricians about this. Too, how much
15 teacher learning is needed to teach in these
16 very different ways.

17 And it is not just about scoring.
18 So we can talk more about that. But the
19 feedback system has to be, to allow teachers
20 to teach in very different ways to help kids
21 explain, learn to explain their reasoning.
22 That kind of discourse goes on in some
23 classrooms. Not in all classrooms.

1 There is a body of work, on what it
2 takes for teachers to have a chance to learn.

3 They need time. They need strategic support.

4 And they need opportunities to try these new
5 ways of teaching. Not just different
6 instructional tasks, out in the context of
7 their own practice.

8 And I am racing a bit, because we
9 are at the end. Build in evaluation and
10 research to the system. And that is the other
11 thing. Besides how they are going to solve
12 the curriculum dilemma that I think you must
13 be asking in the RFP.

14 And lastly, I will say just one or
15 two things about the high school level,
16 because I notice in other hearings it has not
17 been said. People are talking about course-
18 based syllabus. And of course, I am for that,
19 because that is a way to solve this curriculum
20 and assessments; codevelopment.

21 But be aware that we are racing
22 very fast in this country to closure on the
23 fact that there should be just one pathway

1 through high school. And we don't know if
2 that is right.

3 So I hope at least some curricula
4 will look at the phenomenally high level
5 proficiency, especially in European nations,
6 not our U.S. version of vocational ed. But
7 very high level of mathematics that is
8 accomplished when people are involved, let's
9 say, in an electrician's certificate program.

10 And in the white paper, we do talk
11 about that. And give you some references.
12 There are countries that have majority non-
13 academic track percentages of students who
14 outperform us in toto. What does that tell
15 you about that significant number of their
16 students doing career track apprenticeship
17 programs, and they are beating us in
18 mathematics. Thank you.

19 (Applause.)

20 MS. WEISS: Questions for Lorrie?

21 MR. BOND: Yes. Could you go back
22 to your first slide? Let me understand.

23 MS. SHEPARD: My first slide.

1 Okay. Hide your eyes. Don't get dizzy. This
2 one? What did it look like?

3 MR. BOND: Well, I think it is
4 bullet number two there. When you say that
5 you know, you can have a relatively small
6 consortia doing all of these things.

7 MS. SHEPARD: Yes.

8 MR. BOND: Could you talk a little
9 bit more about that? Are you saying, for
10 example, that, you know, northeastern ones
11 should concentrate on just early grades? Or
12 are you talking about just high school or
13 something different?

14 MS. SHEPARD: Yes. I am thinking
15 that it will be very hard. So we all have the
16 NECAP example of four states coming together.

17 They are more like each other than any other
18 four states in the nation, so don't think you
19 can just redo NECAP.

20 And so I am saying that the
21 politics, because we don't want to have happen
22 again what Ed describes has happened
23 nationally. We do not want to have to satisfy

1 neighboring states getting together. They
2 don't have to be even geographically related.

3 That they would water down and do the across
4 the board thing and make the old mistakes.

5 We want them to be ambitious. So I
6 think they will need to be fewer, because they
7 will each want to go through the political
8 processes that we are quite familiar with.

9 And then I am saying both
10 politically and technically, it will be
11 difficult to do all the grades and all the
12 subject areas. So I am saying, that getting
13 it right as a system, and having a more
14 experimental approach will let us do it right,
15 and then generalize.

16 MS. WHALEN: So does that mean, I
17 am sorry. We should think about funding
18 smaller groups of states to do similar types
19 of assessments. Is that what you are
20 proposing? That you have three states doing
21 three eight math then another four states
22 doing three eight math. And then another two
23 states doing --

1 MS. SHEPARD: I think if you tried
2 to do all 50 states in any configuration, you
3 are going to get a tiny tweak better than
4 current practice. So I am assuming fewer
5 states will come with ambitious proposals.
6 And many states will be left out. And I would
7 hope that some would do writing.

8 I am making an important
9 substantive distinction also between the skill
10 areas and content areas. And when you are
11 kind of hoping that it comes in from the
12 field, it would be a mistake if we only got
13 reading and writing. And so we want content
14 areas. And we want skill areas.

15 MS. WHALEN: So how would you
16 ensure, how would you advise us our states to
17 work together to ensure that you then, at the
18 end of this have a comprehensive system
19 without different pieces and different smaller
20 sets of states and expectations. And
21 coherent.

22 MS. WEISS: Can you solve this for
23 us?

1 MS. SHEPARD: Right. They are not
2 going to, you are not going to at the end be
3 able to take all of the pieces and make one
4 national system. What you are going to do is
5 you will end up with every state being able to
6 generalize. So generalization to their whole
7 system could be a criterion in the RFP. So
8 they have to tell you how they have thought
9 about curriculum.

10 And then they have to think about
11 how, if right now we are only trying to do
12 third-grade through eighth- grade writing; how
13 they propose to eventually generalize their
14 agreements and the processes to other subject
15 areas in those grades; and also from the get-
16 go how they planned to extend their continuum
17 into high school.

18 MR. SMITH: Let's go back, Lorrie,
19 because if you had this, in the long run, do
20 you see these different examples coming
21 together to make the system.

22 MS. SHEPARD: I think it is
23 conceivable. It is conceivable that one, no,

1 I don't. Not until the next generation is
2 actually the honest answer, because you do not
3 know enough now to solve the problem of
4 whether or not in October there should be a
5 performance assessment that gets rolled up and
6 scored as part of the external assessment.

7 So that would be an example of
8 something that one consortium might decide
9 that they want to aggregate up high stakes for
10 accountability purposes, end of unit like the
11 benchmarks, but much more substantive. And I
12 want it to count in the overall. And you
13 might get another consortium that says no.

14 We have had too many bad
15 experiences with high stakes not even fitting
16 where the kids are. For example, in Kentucky,
17 they have -- must be on grade level mandate
18 for the roll up portfolio mathematics
19 elements. And half the kids in the class
20 weren't there. So those were inaccurate and
21 wasteful. So you don't want to necessarily
22 mandate that.

23 If two different consortia events

1 solve that end of unit problem in very
2 different ways. You can't just smack them
3 together at the end and say now we have got a
4 national system. But you will have learned
5 all of that. So in the next generation, you
6 can think more coherently. But right now, I
7 am pushing against one national remedy.

8 MR. SMITH: Then let me ask that,
9 you can imagine all of the political issues
10 involved. And what are you proposing? I mean
11 it is -- particularly with NCLB coming down
12 the line, which is a whole different kettle of
13 fish in terms of the people who were going to
14 make final decisions about it.

15 MS. SHEPARD: This slide is for
16 NCLB. That is why the first bullet is there.

17 MR. SMITH: No, I understand. So
18 let me just ask a slightly different question
19 this evening and go back to a question that I
20 asked. In any configuration of this, since
21 there is a common set of standards. You know,
22 you are going to be working within state.

23 Is it enough to create a set of

1 exemplars of the kind of curriculum, not
2 generating the curriculum itself. But a set
3 of exemplars, the kind of curriculum I --
4 areas, performances. Whatever they happen to
5 be. To turn that, those examples into a
6 framework that would then guide the
7 assessments to be both aligned enough and also
8 to be generalized. And also to have some
9 generalization in it.

10 MS. SHEPARD: I think it is
11 possible to take common standards and then
12 this is what we have done badly in the past,
13 is go right to assessments. Right. And to
14 create skeletal frameworks. That has been the
15 problem, that they -- and then it is allowed
16 this wrong kind of alignment.

17 I think we could enrich that level.
18 Now I am trying to do it your way. I am not
19 happy with this. But I am trying to talk
20 through it your way. We could enrich those
21 specifications.

22 And use exemplars that would ward
23 off some of the worst things that Ed talked

1 about. And to be fair then and honest with
2 that way of doing it. It is close to, but not
3 all the way to. What I offered as a remedy, a
4 couple of slides later. And when you said
5 could develop exemplar units. And I do think
6 that those are richer in that they also think
7 through what do I actually have to engage the
8 students in doing.

9 So I think of them like replacement
10 units, where I have thought through what the
11 instruction and the assessment could be. And
12 since -- I would encourage you not to write an
13 RFP that is so narrow that only the first,
14 Mike, is what you allow, because I think that
15 we are not going to really get good at it, and
16 we won't even look like other countries.

17 Hence the Australian example. If
18 we don't take at least that other step to
19 examples. Notice, that is not a full
20 curriculum. But it is at least some where you
21 have thought it all the way through.

22 MR. SMITH: Yes. The earliest
23 Australia standards had lots of really good

1 examples of student work attached to each
2 performance. Each one of the content
3 standards.

4 MS. SHEPARD: Right.

5 MS. WEISS: Okay. Thank you. Over
6 to you, Gary.

7 MR. PHILLIPS: Thank you. So do
8 you set it up, or how does that work?

9 MS. WEISS: Just keep going.

10 MR. PHILLIPS: I am pressing the
11 green arrow. Just keep pressing. Do I get an
12 extra 20 seconds? All right.

13 Thank you very much. I am very
14 pleased to be here, and to talk about what I
15 will call a new generation of comparable state
16 assessments.

17 I want to warn you that I generally
18 disagree with much, maybe most of what my
19 colleagues have just said. I do believe there
20 is a way forward here to reach most of the
21 goals of Race to the Top in a way that is cost
22 effective and in a way that the larger the
23 state consortia, the more likely you can do

1 it.

2 So I would say the goals of this
3 next generation envisioned by Race to the Top
4 cannot in any way be reached by our existing
5 testing paradigm. What we are doing currently
6 in the state assessments, are uncoordinated.
7 They are non comparable. They are non-
8 aggregatable, non-scalable. They are too
9 expensive, and they are too slow.

10 So what I would like you to
11 describe what I will call three pillars of the
12 new assessment system. One of them has to do
13 with common standards. The second one is the
14 use of computer-adaptive tests. And the third
15 one has to do with better measures of growth.

16 Together, these are all three
17 attainable feasible, currently in existence,
18 and they work together as I will show, in just
19 a moment in a large state consortia to give
20 you what you need. First let me talk about
21 the common content standards.

22 The way I would envision this would
23 work is within a state consortium, there would

1 be internationally competitive content
2 standards that lead to high school graduates
3 who are ready for well-paying careers and
4 post-secondary education. This is very
5 similar to the goal of the NGAC assess effort
6 currently.

7 There would be within the
8 consortium, and hopefully a large consortia,
9 again, the larger the consortia the better.
10 Within the consortia there would be a common
11 item bank developed by all the states
12 developed by the teachers across the
13 consortium.

14 There would be a common text
15 blueprint that all of the states would agree
16 to. And each state would administer
17 comparable tests that are equated to a
18 consortia common scale. So there would be a
19 common scale across the consortium.

20 So if there were maybe 30 states,
21 there would be one single scale across that.
22 If we went with the something like is proposed
23 in the NGA, and CCSSO, I think that would also

1 work, where at least 85 percent of each state
2 would cover all the consortia common content
3 standards. And the other 15 percent would be
4 a state supplement. This gives states the
5 opportunity to do things beyond or different
6 from the common standards.

7 Another part of common standards
8 are common performance standards. Common
9 performance standards would be internationally
10 benchmarked, particularly the proficient
11 standard. So for each grade, there would be
12 these standards would be comparable across all
13 consortia and vertically articulated across
14 the grades and on a trajectory that leads to
15 high school career-ready and college-ready
16 performance.

17 In other words, the difficulty of
18 the proficient standard within a consortia and
19 across all consortia would be set in such a
20 way that it is comparable across all states
21 which I will describe in just a moment. In
22 order to accomplish this conventional standard
23 setting, would basically have to be

1 reengineered.

2 Conventional standard setting is --
3 has worked well in small scale assessments and
4 maybe up to a state assessment. But when you
5 get to a larger assessment, it needs to be
6 reengineered.

7 What I would suggest that you do is
8 something like this. Let us say you are using
9 the bookmark procedure. You start by
10 establishing within the consortia an
11 internationally benchmarked proficient
12 standard. You determine that empirically.
13 And once you know what that standard is, I am
14 going to show you an example of that in just a
15 moment.

16 Once you know what that standard
17 is, you then subsequently write the
18 performance level descriptors that describe
19 that standard. And then subsequent to that,
20 all of the other performance level descriptors
21 would be written, let's say for advanced and
22 for basic, or whatever it might be.

23 But you set the standard first.

1 And if all the consortia can agree to that
2 standard, that then gives you a comparable
3 standard across consortia and across states.

4 Adequate yearly progress, reported
5 No Child Left Behind Act, or whatever it is
6 going to be called in the future, would then
7 be based on data that are fair and comparable,
8 and give the federal government comparable
9 information across states, across districts,
10 and across schools. So let me give you just a
11 quick example of this.

12 This is an example of the State of
13 West Virginia. They are currently in the
14 process of revising your standards. In fact,
15 they are intent on raising their standards.
16 And what they wanted to do was to use
17 international benchmarking as a way to help
18 them set those standards.

19 So that when they set them, they
20 don't want to be flying without radar. They
21 want to have some indication about where they
22 should set them, so that they would be
23 internationally competitive. So and this

1 table on the left, what they are using, they
2 are using the bookmark procedure. This is
3 Grade Eight mathematics. And the left are the
4 ordered item pages in the order item booklet
5 for the bookmark procedure.

6 So the items at the top are the
7 hardest, the bottom are the easiest. The
8 second column are the skills scores on their
9 tests, associated with each page in the order
10 item booklet. So what the panelists do is
11 they set a standard based on generally on the
12 content of the test. And then they use impact
13 data within state impact data which is the
14 third column.

15 And right now they have a standard
16 where 53 percent of the students are
17 proficient. And they are asking themselves
18 the question: Is this internationally
19 competitive?

20 Well, when you go on over to the
21 next three columns, after linking the state
22 test to NAEP, you can see how the standard on
23 the state test stacks up on the NAEP scale,

1 just like NCS did the state-mapping study
2 where they took each state standard. This
3 takes each, every possible standard on the
4 state that they might set and puts it on the
5 NAEP scale.

6 And what you see is that if they go
7 with the standard they currently have, they
8 are at the low level of basic on NAEP. Then
9 you could do the same thing. You can link the
10 NAEP to TIMMS and go to the NAEP scale and see
11 where is that standard?

12 And as you can see, it is at the
13 low level of the intermediate. You also can
14 take the international benchmarks on TIMMS and
15 convert those to grades which are in the last
16 column as you can see their standard is like a
17 C. So West Virginia would be asking students
18 to reach a C to be proficient.

19 Well, if they don't like that,
20 which they don't, they are likely to set some
21 other standard. And what they are thinking
22 about is something like this. They have
23 raised the standard. They are now up to about

1 40 percent proficient, 43 percent. They are
2 up in the middle of basic, in the middle of
3 the intermediate and a C plus.

4 Currently, they don't want to go
5 further, because they don't know what the new
6 consortium is going to want to do. So they
7 are waiting for the Race to the Top to play
8 out, and hook up with the consortium of some
9 sort. And once they find out what that
10 standard is, then that is what they would go
11 with.

12 I am thinking it is going to be
13 somewhere like around there. But I don't know
14 what that would be, because as you can see,
15 what would happen in West Virginia, if that is
16 the case, you will have 20 percent proficient.

17 This is a huge political problem in every
18 state.

19 One that ESEA is going to have to
20 deal with, because there needs to be some
21 incentive, financial incentive for a state to
22 take a hit like this if they are going to
23 raise their standards to a point that is

1 comparable across the country. There needs to
2 be some way that they can sell this
3 internally. This is an example.

4 You can imagine that instead of
5 West Virginia, the whole three columns would
6 be a consortia of maybe 30 states. So you
7 would have the same thing; a benchmark
8 procedure for the 30 states, scaled scores,
9 impact data, et cetera. And you would follow
10 the same procedures. That is how you get the
11 international benchmark.

12 Okay. The second pillar, computer-
13 adaptive tests. The current model of one size
14 fits all, the general paper and pencil tests
15 given to all students provides poor
16 measurement for low-achieving students,
17 students with disabilities and language
18 learners and for high-achieving students.
19 Computer-adaptive tests should be used.

20 And I have the word "encouraged"
21 here. I think that is too soft. It is not
22 only, I think the best approach, it is the
23 only approach, particularly as you get into

1 large consortia. They already exist in a
2 number of states; about ten. I have listed
3 ten here. I may have these states wrong,
4 because it changes constantly.

5 But these computer-adaptive tests
6 are already in place in many states. This
7 leads to cost savings, multiple testing
8 opportunities.

9 For example, Oregon has three
10 testing opportunities. Many students will do
11 one in the fall, one in the winter and one in
12 the spring, things like that. It gives
13 immediate feedback. When I say immediately, I
14 don't mean a month later. I mean, immediate.

15 And it leads to shorter tests,
16 because the computer is able to hone in on the
17 students' ability faster than a paper and
18 pencil test can. With computer-adaptive
19 tests, you also can easily create formative
20 assessments and interim assessments.

21 For example, you can easily have a
22 fall, winter and spring interim assessment.
23 The teachers can go to the computer and, for

1 example, say, I would like to get a formative
2 assessment on algebra that is generated for
3 the teacher. It is easy to do. These would
4 be developed.

5 It would be in alignment with a
6 summative assessment. And they would be
7 aligned with the common standards.
8 Constructive response items could also be
9 administered and scored by the computer. In
10 Oregon today, a whole bunch of teachers are
11 sitting down to write constructive response
12 items that will be scored by the computer.

13 And so the technology is developing
14 constantly. Obviously, there will be some
15 states for whom they would want to have some
16 constructive response items and performance
17 items that could not be administered and
18 scored by the computer. And those would be
19 scored independent of the computer by
20 teachers.

21 The more you do it, the slower the
22 process, the more it costs. But you could
23 certainly integrate all of that in with the

1 computer-adaptive testing. Accommodations
2 would obviously be provided, and universal
3 design would be used at all times.

4 This leads to better reliability
5 and more accurate measurement for the high-
6 and low-achieving students. And better
7 measurements for students with disabilities
8 and English language learners. It also gives
9 you better validity, because not only can the
10 selection algorithm be adaptive, it can also
11 be standards based.

12 What I mean by that is, if for a
13 student, if 20 percent of the students items
14 should be algebra, then 20 percent of the
15 items administered to students would be
16 algebra, so it would always meet the test
17 blueprint for every student, whether it is a
18 low-achieving or a high-achieving student. It
19 also gives better validity because at this
20 classroom level, teachers would be required
21 across the classroom, many more items would be
22 administered. And they could cover the entire
23 content domain.

1 So deeper levels of the content
2 domain within that grade would be covered.
3 And those are the -- that is the content that
4 teachers would be held accountable for. And
5 so they would be forced to teach the entire
6 content, not just the items on the test.

7 The third pillar of this new
8 assessment system has to do with better
9 measures of growth. Current growth models,
10 and current growth model issues almost always
11 see negative growth for high achievement
12 students. And you see steep outstanding
13 growth for your low-achieving students. Oh, I
14 am sorry. I didn't do it. Thank you.

15 But these are actually artifacts of
16 paper and pencil tests, because there is a
17 ceiling on a paper and pencil test, the high-
18 achieving students can't go any higher. And
19 so just by chance, they tend to drop down. So
20 you will see negative growth.

21 Low-achieving students can't go any
22 lower and due to chance, they will tend to go
23 higher. And so part of their growth is

1 actually chance, random chance. Both of these
2 phenomena would be eliminated or certainly
3 reduced through computer-adaptive testing,
4 because the computer-adaptive tests has a very
5 low floor or no floor and a very high ceiling.

6 You can go up and get the high
7 achievement students and can go down and
8 adequately assess the low achievement
9 students. This would require a vertical
10 scale. If you don't want to measure growth,
11 you don't need a vertical scale. But if you
12 do want to measure growth, I would argue that
13 you need a vertical scale.

14 And I would also argue that people
15 that do growth models without a vertical scale
16 is bogus. And you shouldn't do it. So what
17 you would need here is a vertical scale to do
18 a good growth model.

19 Value added indices for teachers
20 and principals would also be based on this
21 vertical scale, and it would yield comparable
22 data and more fair and accurate comparisons
23 for teachers across the consortia. This of

1 course would require a statewide longitudinal
2 data system.

3 You are going to be spending part
4 of \$40 billion to make sure that exists on all
5 the states, so you should have that in the
6 near future. A good vertical scale could not
7 be developed without that. But again, this is
8 another good example of technology that is now
9 in place that was not in place, that will be
10 in place in the future that was not in place
11 in the past. And this makes vertical scales
12 possible.

13 Growth models are themselves
14 inherently unreliable, less reliable than
15 status measures. That is because growth is
16 the vertical progress between two status
17 measures. And since status measures are
18 themselves unreliable, the difference between
19 them is even more unreliable.

20 The beauty of computer-adaptive
21 testing is that the status measures would be
22 more reliable, and therefore, as a logical
23 consequence, the growth measures will be more

1 reliable. So a nice byproduct of computer-
2 adaptive testing would be more reliable growth
3 measures. Okay.

4 What are the overall benefits of
5 this new design? First, it implements the
6 vision of Race to the Top. We have higher
7 quality assessments based on fewer, clearer
8 and higher standards. It improves No Child
9 Left Behind by correcting two of its most
10 fundamental problems.

11 We have too many content standards
12 and too many performance standards. This
13 design is beautifully scalable to large
14 numbers of states, and it takes advantage of
15 innovation and technology. You will get a
16 better measurement of a wider range of
17 students in the general population.

18 If I had more time I would also
19 explain how this is implementable. And the
20 alternate assessments, the 1 percent
21 population. And this would eliminate the need
22 for the 2 percent, modified 2 percent. The
23 reason why there is a modified 2 percent

1 population is, that everybody is using paper
2 and pencil tests that can't go down and
3 adequately measure those students in that
4 population.

5 So we are in the process of
6 developing a whole new assessment system to
7 try to get down there. But the computer-
8 adaptive testing does go down there. It goes
9 down and gets all those students and gives the
10 best measurement for them. And it goes up to
11 the top and gets your high-achieving students
12 and gets a good measurement of those.

13 This system is also feasible and it
14 meets all of the professional and technical
15 standards of the APA, MCME and AERA. It is
16 affordable. And in the long run, I believe it
17 would cost about half as much as traditional
18 pencil and paper tests.

19 It benefits the federal government
20 because it gives them what they want. It
21 gives them comparable data for states,
22 districts and schools. And it benefits the
23 states, and gives them cheaper, faster and

1 better assessments and gives them some local
2 flexibility. Thank you very much.

3 (Applause.)

4 MR. BOND: Gary, this is, you know,
5 a beautiful vision. And I am actually happy
6 to see this, because part of my presentation
7 said that we might be able to agree on all of
8 these things. And then I hear Ed and I hear
9 Lorrie, and I said, well, maybe I ought to
10 strike those slides.

11 I just got a report from NCES that
12 talked about how different the states were in
13 terms of their proficiencies on NAEP. And I
14 mean it was some huge differences. And I am
15 just wondering if you would respond to that,
16 given what you just said.

17 MR. PHILLIPS: Sure this is in fact
18 is intended to fix that problem. That is
19 correct. In the current No Child Left Behind,
20 the states with the highest percent proficient
21 are the lowest achieving states on NAEP. So
22 one of them, at least one of them is wrong.

23 MR. BOND: Yes. I mean South

1 Carolina was really off the charts.

2 MR. PHILLIPS: Well, the standard
3 for South Carolina is off the charts.
4 Performance was not.

5 MR. BOND: Yes. Performance was
6 not.

7 MR. PHILLIPS: Right. So what you
8 need here is a situation where you have
9 comparable proficient standards across the
10 states, this will be difficult because the
11 low-achieving states that currently have a lot
12 of proficiency are not going to have that in
13 the future. So this will have to be dealt
14 with in ESEA.

15 But this system would fix that.
16 Every state would buy into the same proficient
17 level. It would be internationally
18 benchmarked. That would have to be decided in
19 the future.

20 If I were asked to give advice on
21 that, I would say it should be paid to the
22 NAEP proficient or the TIMMS high level. And
23 that makes those standards something that the

1 states have never had before. It makes it
2 internationally competitive.

3 By the way, I think this is a
4 quintessential role for the federal government
5 to do this, because the federal government has
6 historically had two main roles in education.

7 Equity and providing good information on the
8 condition of education.

9 What this does, is essentially
10 removes what I would consider to be an
11 equitable situation where the students in the
12 states that say they have a large number of
13 proficiency by setting low standards, it is
14 not giving the students in those states a fair
15 shot at the same kind of expectations that you
16 have like in a state like Massachusetts. So
17 it is an appropriate role for the federal
18 government to create a system that would fix
19 that, and to do a better job of reporting on
20 the condition of education in the country.

21 MR. LINQUANTI: Gary, I have a
22 follow-up question to what Lloyd was asking.
23 And it's, what you are saying has been

1 reinforced by a number of my colleagues, Eddis
2 Culnets and others, who are saying that
3 computer-adaptive testing really has a
4 potential we are not tapping.

5 And one of the things that I bump
6 up against in my empirical experience is where
7 schools are. And so I am thinking the
8 feasibility question about can we implement
9 this on a large scale, when I see schools that
10 still don't have access to the internet. What
11 are the technological feasibility issues that
12 we would have to address?

13 MR. PHILLIPS: Okay. That is a
14 great question, and I have two answers. One
15 is, states and districts of schools that don't
16 have computers should have them. This is a
17 great leverage to get them, partially paid for
18 by the federal government hopefully. That
19 would trickle down into the schools.

20 Secondly, the nice thing about
21 having multiple testing opportunities, for
22 example, I will use Oregon as an example. And
23 by the way, Hawaii is about to do the same

1 thing. With multiple testing opportunities,
2 you don't need to have a computer. Every
3 student doesn't need to have a computer,
4 because you are using the same computers over
5 and over again.

6 So with just a very few computers
7 in the school, you can test the whole school.

8 So it is kind of a mistake to think you have
9 to have a computer for every student. All you
10 need is some computers, internet access,
11 broadband, things like that. And schools that
12 don't have that should have that. This is
13 great leverage to get it.

14 MS. WEISS: Mike.

15 MR. SMITH: It strikes me that you
16 folks are addressing different questions. And
17 Lorrie, it is a different question, I think.
18 Just to put it in one way, I think you are
19 trying to address issues of construct
20 validity.

21 That is some sort of much richer
22 way of thinking about assessments and of tying
23 it to the actual content and skills that

1 students obtain. And that strikes me as
2 somewhat different from what you are
3 addressing, which are technical issues which
4 are important. And I think the issue around
5 computer-adaptive assessments is really
6 important for us to dig into.

7 And it would be useful to have
8 comparable levels of proficiency if we go that
9 direction. On the other hand, the idea of
10 giving a set of assessments that have a tight
11 enough link to, conceptual link to the
12 curriculum and to the nature of what is going
13 on in the classroom.

14 And to the -- and in addition, the
15 idea that you have got some sort of adaptive
16 instruction. You have got a way of thinking
17 about a teacher and student relationship
18 through formative assessment, through tight
19 formative assessment at the point of
20 instruction, those things strike me as being
21 really critical for us to move forward in a
22 big step. And I see what you are suggesting
23 is a little step and maybe not even in the

1 direction that we want to end up going in.
2 But there is a big step.

3 Greed -- it is going to be much
4 harder to get to, and it may be more
5 expensive. I think we need to have built in
6 continuous improvement as you go through the
7 process that Lorrie and Ed have suggested.
8 But I worry about taking the next step down
9 the same old path. And that is what I think
10 you are doing. Can you comment on that?

11 MR. PHILLIPS: Yes, sure.

12 MR. SMITH: Gary and I are old
13 friends.

14 MR. PHILLIPS: No. That is a great
15 question.

16 MR. SMITH: I am not being, as you
17 know, I am not being critical. We have been
18 through these discussions before.

19 MR. PHILLIPS: No. I think that
20 that is a great point. All of the issues
21 about curriculum alignment that the need for
22 performance-type assessments, the need to have
23 high-end, challenging-content standards. The

1 need to have R&D, research innovation. I see
2 all those, I see that as going on
3 simultaneously with the development of a
4 computer-adaptive test.

5 I don't see those as contradictory,
6 or as -- and I certainly don't think we should
7 wait around for another ten years and work
8 through all that when we don't need to. So I
9 think all of those issues of validity, and
10 alignment and all of the things that Lorrie
11 and Ed mentioned are all important and need to
12 be dealt with. But you don't need to put
13 everything on hold while those are being dealt
14 with.

15 Particularly, I mean what is going
16 to happen I think is, if the federal
17 government does nothing, the states themselves
18 are going to go down this direction one by one
19 as more and more technology is developed and
20 it becomes more and more feasible and cheaper.

21 And you are going to be in a situation where
22 the states are ahead of you. I am just
23 suggesting you might want to get out in from

1 of this.

2 MS. WEISS: Lorrie.

3 MS. SHEPARD: I think that this
4 exchange is very important, and we should save
5 for the round table the relationship between
6 adaptive testing, which I would like to speak
7 for, and the substantive and structural issues
8 that are behind your questions. It is too big
9 for me to get in one --

10 MS. WEISS: Yes. I wrote it down
11 as a question for this afternoon.

12 MS. SHEPARD: So let me just say,
13 though, I can make a clarifying comment in a
14 half a minute, having to do with the cut
15 scores for proficiency. That is a completely
16 different issue.

17 The white paper that I referenced
18 in the second slide addresses it specifically.

19 It argues against using proficiency cut
20 scores as the primary way to reporting because
21 it so distorts. And there are simple ways.

22 And Gary is one of the best
23 scholars. But Henry Brown is another. And

1 Don McLaughlin have all done the equating that
2 let us see so clearly how crazy the different
3 cut scores are. That is a different problem
4 and should be addressed.

5 I put it in the same box as the
6 whole speech we could give about AYP, all of
7 which are not on the table today. But I do
8 think they have to be engaged in this round by
9 the states.

10 MR. PHILLIPS: Yes. We need more
11 meetings.

12 MS. WHALEN: Just a quick
13 clarifying question. In your vision of the
14 assessment system and the use of adaptive
15 tests, would you allow for off grade level
16 testing as well, or just within grade level.

17 MR. PHILLIPS: I think that is
18 optional. It could be done within grade
19 level, or it could be done with off grade
20 level. That is another big issue. I think
21 going either way doesn't have anything to do
22 with the use of computer-adaptive testing.

23 MS. WEISS: And I wrote that down

1 for a conversation after at the round table
2 also.

3 MS. MELENDEZ: Just really quickly,
4 you mentioned English language learners. How
5 does that work in terms of the --

6 MR. PHILLIPS: Well, one of the --
7 again, just like students with disabilities,
8 not those in the 1 percent but those that are
9 not in the 1 percent, traditional paper and
10 pencil tests are too difficult for them, or
11 have language barriers. And so it makes it
12 difficult to go down to where they are, to get
13 good accurate measures. As you build in
14 accommodations and that sort of thing, in
15 computer-adaptive tests, you get better and
16 better measurements of students that are a
17 little low-achieving.

18 MS. WEISS: Okay. I am going to
19 move us on to Lloyd.

20 MR. BOND: Okay. I would like to
21 also thank everyone for inviting me here.
22 This is a great opportunity for me to get some
23 things off of my chest.

1 Let me just start by saying that I
2 think everyone on this panel could write a fat
3 book about our charge here. But we are forced
4 to be reminded of Bill Strunk's quip that in
5 every fat book, there is a thin book trying to
6 get out. And so these are our thin books.

7 I had hoped that after looking over
8 the Boston meeting and the Atlanta meeting I
9 thought that there was going to be easy issues
10 about which we could agree. I am thinking now
11 that I may want to skip over these first few
12 slides. But I will put them out anyway.

13 We will probably be able to agree
14 on intended uses, on contents, formats, test
15 administration. I thought that was so before.

16 I don't think so now. But to who --
17 prestigious for validating this thing are
18 probably straightforward, and I don't see any
19 problem there, as well as that reporting to
20 the public. So I am going to skip over these.

21 You can ask me more questions about them
22 afterwards.

23 I would like to spend my time

1 talking about what I think are unresolved
2 issues here. And we are going to start
3 with -- excuse me -- the broad and shallow
4 versus the narrow and deep problem. This is a
5 huge, huge issue here. If you look at any
6 state's curriculum, and in fact, any textbook,
7 it is so much in there for any teacher to try
8 to attempt to even cover in one semester is
9 absolutely absurd.

10 Of the kinds of things we expect
11 teachers to do in four or five months. And I
12 will come back to each one of these in a bit
13 more detail. The consequential validity
14 issue, that is, teaching to the test; the
15 concepts of college and "work readiness" and
16 finally what I call familiar and aspirational
17 assessment.

18 First, the broad and shallow versus
19 narrow and deep. There is this very delicate
20 balance between learning a lot of things
21 marginally and learning a modest number of
22 things very well. And I think we as a nation
23 have erred seriously on the former.

1 It is a fact that one of the
2 distinguishing curricular features of
3 countries that consistently do well on TIMMS
4 is that they err on the side of narrow and
5 deep almost consistently throughout. Here are
6 sort of my summary thoughts on how the
7 architecture of this assessment should
8 proceed. One from the assessment of what I
9 call factoids. Basic declarative knowledge
10 through starting in the very early grades.

11 I mean, kids have to know facts
12 first. They can't reason without a
13 fundamental declarative knowledge base. So we
14 sort of start there, and we end up at this
15 ideal of problem solving and higher order
16 thinking.

17 And I am going to talk a bit more
18 about what I call the assessment of
19 understanding, which I think has been totally
20 neglected by just about everybody: the
21 teaching to the test problem.

22 I have written a very brief essay
23 on the Carnegie website. It is called

1 Carnegie Perspectives. And I invite you to
2 read it. There is no question but what, if
3 and when these assessments are used for
4 accountability purposes, and they will be, the
5 teachers are going to teach to the test. And
6 so what we have to do is to make sure that the
7 test is worth teaching to.

8 And that is no mean task. I am not
9 against, I mean, I think teaching to the test
10 has acquired a bad name. And my argument in
11 the essay is that we can in fact construct
12 tests that teachers should teach to.

13 I have had, and I continue to have
14 deep questions about what we mean by college
15 and work readiness. And here is my
16 fundamental problem. Ready for majoring in
17 phys ed at Podunk U., versus ready for
18 majoring in EE at Cal Tech. Ready for a
19 cashier at McDonald. Ready for an office
20 manager trainee at Merrill Lynch.

21 It is doubtful whether a single
22 test or even a series of tests can serve these
23 multiple purposes. So I think we have to

1 think fairly carefully about this notion of
2 "work readiness." Here is one of my favorite
3 sayings, and I said it in many different
4 contexts.

5 We don't know how yet to measure
6 all of the important things. So that what we
7 can measure becomes all important. We are
8 sort of stuck with that fact. And I think
9 that the RttT assessments should encourage and
10 support the consortia that move the
11 professionals forward. We are stuck still
12 with you know, quick recall, fast little
13 multiple choice tests and even some of our
14 extended response exercises don't get at what
15 I call a student understanding.

16 Let me give you a couple of quick
17 anecdotes. There was a famous, first -- I was
18 on a dissertation committee for a math ed
19 student. And she wanted to see how whether
20 grad students in math and science ed
21 understood a popular procedure for extracting
22 the square roots by hand. And very few could
23 explain why the procedure worked.

1 And she was intrigued by this. So
2 she said, well, let's see if they can explain
3 why long division works. And they couldn't
4 explain that either. Now they were adept at
5 these procedures. But they didn't quite
6 understand why they worked.

7 A second is a series of studies
8 conducted at Johns Hopkins of first-year
9 physics students there. Okay. Now, these are
10 sharp kids, okay. They made the cut at Johns
11 Hopkins. They were adept at the equations
12 that explain the movement of moving bodies,
13 but they had no idea of what the implications
14 of these were, for the actual world.

15 For example, they were asked to
16 draw the trajectory of a ball cut from a
17 pendulum that their equations implied. They
18 had that ball doing some of the craziest
19 things you can possibly imagine. It was part
20 this way, and shoot straight down. They cut
21 the ball at the very top of the pendulum and
22 it would shoot out this way.

23 Some of the craziest notions about

1 how the world works, but they could solve all
2 of the equations. I mean, they could do the
3 problems in the back of the book, after each
4 particular chapter. So it got me to thinking
5 that one area where we really do need a lot of
6 work is the measurement of understanding.

7 Again, I have rewritten a piece on
8 this in the Carnegie website. I invite you to
9 read it. But the point is, that trying to
10 assess understanding will also change the way
11 teachers teach. If they teach for
12 regurgitation, which many of them do, students
13 will forget. If they teach for understanding,
14 students will remember.

15 Some final thoughts. Beware of the
16 assessment tail wagging the content dog. This
17 is a huge problem. And we must always let
18 content dominate and let the test follow.
19 Granted, sometimes we cannot measure many of
20 the things we would like to. But we should
21 keep this simple principle in mind.

22 We should not reinvent the wheel.
23 We should encourage consortia to take

1 advantage of what we already know and what has
2 already been done. And finally, we should
3 encourage the involvement of high ed faculty
4 and industry personnel in all of the high
5 school assessments. Thank you.

6 (Applause.)

7 MS. WEISS: Questions?

8 MR. PELLEGRINO: Lloyd, I just want
9 to clarify one thing. When you talk about
10 teaching facts, you are not talking about
11 teaching them separate from being able to
12 reason with the facts at the same time. Or
13 are you just taking a strict few, of just the
14 facts first?

15 MR. BOND: Well, you know, I have
16 observed algebra classes, third grade
17 arithmetic assessments and even at the high
18 school level, they teach facts. They don't
19 teach how to solve for problems. It ends up
20 being here is how to do this, here is how to
21 do that. Here is how to do that.

22 And what the kids end up with is a
23 list of unusable inert declarative knowledge

1 facts that are totally inaccessible during
2 problem solving. So I mean, what I am saying
3 is, at the early grades, kids have to know
4 certain things.

5 I mean, they have to know what a
6 fraction is, perfect example. Lei Ping Ma at
7 the Carnegie Foundation located American
8 fourth-, fifth- and sixth-grade science -- I
9 mean, general ed K through five teachers who
10 could not explain the difference between
11 dividing by two and dividing by one half.
12 They didn't want a verbal level what that is,
13 but they couldn't explain why. And having
14 them teach students fractions.

15 So what they end up doing is in
16 fact teaching facts rather than teaching
17 reasoning. I am not sure at the second, third
18 and fourth grades, Jim, how much reasoning
19 kids can actually do, but they still need to
20 have some basic declarative knowledge and base
21 in order to even start.

22 MR. PELLEGRINO: I think this is a
23 bigger topic for later.

1 MR. SMITH: Yes. I think that is
2 right. I think it ought to be in the round
3 table. This, the issue of understanding. I
4 thought, Jim, you were asking a question of do
5 you get assessed on a series of facts, and
6 then do you get assessed right after that
7 somehow on your understanding of why this
8 phenomenon is happening.

9 I don't know how well we can assess
10 whether or not a student understands that. I
11 suppose we could create items that do that.
12 But I do think it is for later. And it should
13 go along with things. Lloyd has raised at
14 least five different issues in my mind on
15 this.

16 The idea of learning to learn goes
17 in part to your question of what is college
18 ready, and what is career ready. I mean, if a
19 student understands that how to learn from a
20 bunch of facts and how to take the next step
21 of how to generalize from a body of
22 information, that strikes me as being more
23 predictive probably of how well they will do

1 in college and work than a set of kind of
2 random facts about arithmetic and verbal
3 skills that we typically give.

4 But this again, I think, is a
5 larger issue. So these are really important
6 issues, I think. And it may require a
7 different kind of test item or of approach
8 than you have seen in the past.

9 MR. LINQUANTI: The other thing
10 that I would add for possible discussion later
11 on, Lloyd, your example about what readiness
12 actually gets implemented as. And I think it
13 really is a question about the expectations as
14 expressed, versus the expectations as
15 implemented. And it is an equity issue.

16 And I think we need to think hard
17 about what this means, and how it gets
18 implemented, because you see it in things like
19 course sequences in high school where folks
20 may be saying, oh yes, these are college-ready
21 course sequences. And yet there are notable
22 differences in their de facto tracks working
23 underneath there.

1 MR. BOND: But the college faculty
2 of -- I was involved in a Carnegie Foundation
3 study of why it is that community college
4 students do so -- I mean, 80 percent need some
5 remedial instruction in math. And it turns
6 out that what high school teachers think
7 college readiness is, and what higher ed
8 faculty think it is, are completely different
9 things. And so we have got to get these two
10 communities together in this effort.

11 MS. WEISS: Lorrie.

12 MS. SHEPARD: Mark and Joanne in
13 preparing for this asked me to attend to early
14 childhood issues. And it is not featured in
15 my presentation because we were also asked to
16 talk about summative tests, which I think then
17 precludes talking about prior to grade eight.

18 But in this context, let me say a
19 couple of things on behalf of young people.
20 They can reason. And how they are taught
21 about reasoning is an important curricular
22 issue. And when your two-year-old asks you
23 over and over and over again why, and you tell

1 them why, you are teaching them cause and
2 effect.

3 And we can -- with very young
4 children teach them to explain their
5 reasoning. And make it a matter of how we --
6 that is that one slide about classroom
7 discourse. If we go deep into what is behind
8 that research is, you can teach kids to have
9 to explain.

10 And kids are disadvantaged by their
11 schooling when in some classrooms that is
12 normative practice, and in other classrooms it
13 is not. So this is very much what we are
14 talking about, about curriculum.

15 MS. WEISS: Ed.

16 MR. HAERTEL: You talked about
17 broad and shallow versus narrow and deep.
18 There is a delicate balance. Sort of framing
19 it as a continuum where we need to find the
20 right point on that continuum.

21 And I am wondering whether it is
22 better thought of as sort of a qualitative
23 divide. Different conceptions of what the

1 curriculum is supposed to be about and what
2 the important outcomes of schooling are. And
3 you sort of -- being a sort of analytical
4 fellow, you are sort of suggesting we should
5 try to split the difference and set it for
6 everybody.

7 MR. BOND: I -- no.

8 MR. HAERTEL: Is that realistic?

9 MR. BOND: No. I accept your
10 modification.

11 MS. WEISS: Ann?

12 MS. WHALEN: You have raised a
13 bunch of really challenging conceptual ideas
14 about what a test should be able to do, and
15 how it drives teaching and learning in the
16 classroom. What advice do you have for us at
17 the Department in terms of formulating the
18 actual notice and application to get at these
19 issues in terms of questions or expectations?

20 MR. BOND: The only thing I would
21 suggest is that you favor or you encourage --
22 I mean, you make that part of the RFP; that
23 they consider new forms of assessments, and

1 new things to assess as part of the RFP.

2 MS. WHALEN: So you would suggest
3 that we ask for evidence from states in
4 developing their assessment system that they
5 show us how they are going to assess these
6 different types of knowledge?

7 MR. BOND: Yes.

8 MS. WHALEN: How they are going to
9 weave that into whatever system they create?

10 MR. BOND: Not just assessment as
11 usual. You know, what are you doing to get at
12 some of the competencies that we all know are
13 important but that are never assessed. Like
14 understanding.

15 MS. WHALEN: Can you kind of
16 articulate any examples that you would see for
17 how states would get at that?

18 MR. BOND: Yes. My essay goes into
19 five or six different way that I think we
20 should approach this. And I will just use as
21 a teaser for you.

22 MS. WHALEN: Do your homework?

23 MR. BOND: Yes.

1 MS. WEISS: Okay. I think we are
2 going to take a break for lunch. We will
3 reconvene here at 1:15, and we will start
4 right on time. So do get back by 1:15.

5 And remember, if you have got
6 questions that have come up so far, don't
7 hesitate to drop them at the table on your way
8 out, and we will gather them up and use them
9 for our conversation this afternoon. So
10 thanks, and we will see you back here at 1:15.

11 (Whereupon, the morning session was
12 concluded at 12:15 p.m.)

13 MS. WEISS: Hello. Everybody come
14 back and take your seats. We are ready to get
15 started, so we are going to reconvene with
16 Robert Linquanti.

17 Robert, do you have this?

18 MR. LINQUANTI: Yes, I do.

19 MS. WEISS: Great.

20 MR. LINQUANTI: I have the --

21 MS. WEISS: The clicker? Okay.
22 Terrific. Thank you all for coming back and
23 getting settled. And let me turn it over to

1 you, Robert.

2 MR. LINGUANTI: Okay. Thank you,
3 Joanne. First of all, thank you for coming
4 back from lunch. We really appreciate that.
5 And thank you for having me part of this
6 panel.

7 When I was asked to join this
8 panel, I had a little bit of cognitive
9 dissonance. I said, now why would they want
10 me on this panel, given the other folks that
11 are here. And I think, this is my theory is
12 that because I actually am not a
13 psychometrician, and I am not necessarily -- I
14 don't consider myself an assessment expert.

15 I do a lot of work with districts
16 and states around accountability and
17 evaluation issues related to the assessments
18 that are used, and a lot of technical
19 assistance with states on those things. And I
20 figured it was because I really focus a lot on
21 English language learners, which is obviously
22 a very key population for a number of the
23 states, and particular some states in the room

1 here today. And so that is the role I am
2 going to assume.

3 I know there is an EL panel
4 specifically tomorrow. In fact, I am on it.
5 So consider this to be a kind of foreshadowing
6 if not a more than half of the presentation
7 that I will make tomorrow.

8 But I thought this would be a very
9 nice way to look at this particular population
10 and it is very nice to come after Ed, Lorrie,
11 Gary and Lloyd and before Jim, because I can
12 just build a lot on what all of these people
13 have said already. And anything that I say
14 that is wrong or dumb will be quickly
15 forgotten when Jim speaks.

16 So I think there is a lot of
17 strategy here in this spot . So particularly
18 the question that I am asking is very
19 specific. How will English learners
20 participate in this? And I think it is a
21 question that each of us needs to think about,
22 particularly in the states as you think about
23 you know, how might we include English

1 learners in this?

2 And so my goals for this session
3 just are to, and I am running this in parallel
4 on my laptop. But it doesn't advance when I
5 hit the advance clicker. Are to touch on a
6 few key aspects of English learner status
7 relevant to the topic of developing a
8 comprehensive assessment program. Then
9 indicate how this assessment program offers
10 what I think are major opportunities for
11 improving assessments and instruction for ELs
12 if we go about it the right ways.

13 And it is not one way. There are
14 several ways to go about this that I think.
15 And finally, suggest what it will take to get
16 there. That is very ambitious for 20 minutes.

17 I am going to skip and jump and touch on
18 things rather lightly.

19 I am sure we will get into more
20 depth tomorrow. Since this is the general
21 assessment session, I wanted to make sure that
22 those of you who do not stay for tomorrow at
23 least had a chance to taste some of these

1 issues and ideas. And hopefully, we can weave
2 them into the broader discussion this
3 afternoon.

4 So first question I want to ask
5 are, which ELs are we talking about? When we
6 talk about English learners, we need to recall
7 that this is not a monolithic group.

8 In fact, there is more variation
9 within the EL group than there is compared to
10 the non-EL group. And by that I mean, we have
11 English learners that are born in the U.S. and
12 are there their entire lives, versus the kids
13 who came when they were early adolescents,
14 versus the more recent arrived immigrant, the
15 1.5 generation, so called.

16 They vary considerably by their
17 time in U.S. school, and also their age and
18 grade on entry. They also vary by whether
19 they come to U.S. schools literate in their
20 first language or not.

21 And that has a lot to do with their
22 prior formal schooling, as well as their
23 socioeconomic status from their home

1 countries. Many times, we can have kids that
2 are free reduced lunch in the U.S., but they
3 are coming from their home country and were
4 actually middle class or higher in their home
5 country. So they may come very well prepared
6 academically in their home languages, but that
7 may not be reflected in their current economic
8 status.

9 Certainly, in terms of school
10 attendance and consistency, this term there,
11 SIFE, students with interrupted formal -- I am
12 sorry. I'm not doing this. No. I have to
13 close this now, because I am obviously not
14 keeping up.

15 Students with interrupted formal
16 education, so there is a real difference in
17 terms of school attendance and the consistency
18 of attending. When we look at years in U.S.
19 school, we need to think about, is that
20 continuous schooling, or are those kids that
21 have come and then gone for four months, and
22 then come back again.

23 And then finally, the most obvious

1 thing, which I try to put last, or next to
2 last are when we talk about a kid who is an
3 English language learner, besides that being
4 somewhat of a deficit view, what they lack,
5 versus what they come with. Which is what
6 these other points try to make; also there are
7 varying levels of English language
8 proficiency.

9 There are kids that are virtually
10 non-English speaking. There are kids that are
11 at intermediate levels. And then there are
12 kids that are actually at quite advanced
13 levels.

14 And we also need to think about
15 what that overall categorization may mask in
16 terms of the different subskills. They may be
17 very good in listening and speaking but very
18 weak in reading and writing. So we need to
19 sort of pull apart those definitions and look
20 at how that composite score is defined.

21 And then finally, a good colleague
22 of mine, a former colleague, Guillermo Solano-
23 Flores, who is here at U.C. Boulder talks

1 often about the sociocultural and dialectical
2 differences. If we think that Spanish-
3 speaking ELs are all the same in terms of
4 their language demands or even giving out one
5 content assessments, primary language
6 assessments.

7 We actually find quite a bit of
8 variety. The Spanish of Mexico is very
9 different from the Spanish of Guatemala or
10 Venezuela. So we really need to think about
11 these things.

12 How does one get into and out of
13 the EL category again. This is more
14 foreground. Now I want us to get a clear
15 sense of who these kids are, because this has
16 huge implications for how we will include them
17 or not in these assessment systems we are
18 talking about. So clearly, a kid gets in.

19 There is a home language survey
20 that identifies language minority status, and
21 triggers an initial ELP assessment. So right
22 from the beginning, what gets them into the
23 category is a survey questionnaire. And then

1 an assessment.

2 And that assessment will identify
3 whether you are an English learner, versus
4 whether you have enough English to be
5 considered fluent proficient. And these are
6 probably more true bilinguals or English
7 dominant.

8 What gets you into the EL status,
9 and this is critical, it may be very different
10 from what gets you out of it. There are
11 several states. And many of them with large
12 EL populations, that will have kids come into
13 the EL category based on a language
14 proficiency test but exit the EL category
15 based on a language proficiency test as well
16 as certain levels of scores, threshold scores,
17 in academic achievement tests, academic
18 content assessments.

19 So we have an issue here, where EL
20 status may be defined in part by a content
21 achievement test, let alone the noncognitive
22 measures in states that have local definitions
23 allowed. They may include grades or

1 attendance, or homework hand-in. I mean a lot
2 of things get rolled up into that. Parent
3 consultation. So we need to keep in mind the
4 construct is very complex and varying.

5 I am going to skip this. This is a
6 classic example of a schematic that is really
7 not good for large audiences.

8 Basically, what it is trying to
9 show you is, that once you get into the EL
10 category, you either will meet the criteria to
11 exit that category and become a former English
12 learner. Or you will stay in that category
13 because you are not meeting one or more of the
14 criteria and you run the risk and I say risk,
15 because it is a risk, educationally, of
16 becoming a long term English learner. And we
17 need to talk about that when we are thinking
18 about our assessment systems.

19 Because if assessments are in part
20 defining this population, then it is our
21 obligation to monitor the progress on each of
22 those assessments for these students. That is
23 the kind of thing that this program is

1 suggesting is important.

2 And it links to of course,
3 formative assessments but also the interim
4 assessments and ultimately, the summative. We
5 need to think. This has huge implications for
6 these kids.

7 If they are EL and hitting the
8 secondary level, they have a very high chance
9 of being trapped into lower expectation
10 curricula and lower level materials. And that
11 is all too often a dead end, or a recipe for
12 dropping out. So there are huge implications
13 here.

14 So just to touch lightly on this.
15 There is a skimming effect. The most
16 successful ELs will typically exit the EL
17 cohort. That leaves the lower performing ELs
18 and the newly arrived ELs.

19 Okay. So when your people say, we
20 have a chronic problem with these ELs. They
21 just always do poorly. But look at these
22 reclassified fluent English proficient kids.
23 They do really well. So what we need to do is

1 just reclassify more kids.

2 Well, there is a definitional issue
3 there, right. In that part of what gets you
4 reclassified is a certain level of performance
5 on academic content assessments, typically.
6 And even if it is just the LP assessments, you
7 know a lot of those are pegged to certain
8 performance levels in academics.

9 So we have a definitional issue
10 here. Right, a sort of a skimming bias.
11 Those remaining are not meeting certain
12 criteria. Which criteria? The better
13 performing by definition exited ELs may still
14 have ongoing linguistic and academic needs.

15 And so what this suggests is
16 although we have binary categories and funding
17 based on them, we actually have a continuum of
18 progress that kids make. A continuum of,
19 really, expectations or progress expectations
20 which we should be articulating clearly. And
21 then the needs and supports that occur while
22 they are English learners, but actually
23 continue after they are English learner.

1 Because let me tell you, the
2 language demands of curriculum and of tasks do
3 not get easier as you go up the grades. They
4 get harder. And kids reclassified or exited
5 at the third grade or fourth grade may be
6 having significant academic language demands
7 and issues at the sixth to seventh grades.

8 We need to have a system that is
9 sensitive enough to recognize that, and not
10 constrained by definitional categories. Okay.

11 You meet the kids where they are, and serve
12 them where they are needed, both
13 linguistically and academically. So I hit the
14 wrong button there.

15 So let me just remind you again.
16 This is all background to where we are going
17 here, with this. Which is, we are on the
18 hook. Educators are on the hook for two
19 things. To provide meaningful access to grade
20 level academic content via appropriate
21 instruction. That is a mouthful. And each of
22 those phrases is important.

23 The second thing we are on the hook

1 for is to develop kids academic English
2 language proficiency. And I deliberately put
3 that order, because so often, non-specialists
4 tend to think of number two as being first and
5 prerequisite before number one can actually
6 occur. And that is a mistake. In fact, this
7 needs to occur simultaneously and non-
8 sequentially.

9 It may seem contradictory, but
10 there are ways, methods, instructional methods
11 that we are working on that we need to make
12 sure are in place to ensure, kids you -- even
13 at the lowest levels of language proficiency
14 have meaningful access to cognitively
15 challenging grade level content. It is being
16 done now, and in fact, it can be assessed.
17 And that is where this competition has to get
18 it right, because there is 5.1 million of
19 these kids out there. And they cluster most
20 typically in high poverty schools.

21 And when we think about the other
22 aspects of RTT Race to the Top, and the focus
23 on the lowest achieving schools, we have some

1 serious equity issues here. And we need to
2 make sure that we address them, through all of
3 the vehicles that this administration is
4 putting out. Fortunately, that has the
5 resources to put out. Dan hitting that wrong
6 button. All right. So academic language
7 proficiency. Let me just touch on this
8 briefly.

9 I had colleagues who said to me,
10 Look, you are talking to a general assistant
11 office. Don't lead with language. They will
12 turn off. They are going to say, all right.
13 These are the language people. They are going
14 to talk about ESL, and English language arts.

15 I do math and science. I am just going to
16 check my BlackBerry. So we are watching.

17 But the point is, it is really hard
18 not to talk about languages. That is such an
19 important defining characteristic of these
20 students. It is not the only defining
21 characteristic. ESL or ELD with an academic
22 focus is necessary but not sufficient for
23 these kids to succeed academically.

1 Many if not most of the academic
2 tests are mediated by language across the
3 disciplines. And what that means, obviously,
4 is that every teacher must model and teach
5 using the language of their discipline. And
6 they also need to make sure that language is
7 required for completing grade level tasks in
8 classrooms.

9 When you go into most classrooms,
10 and I go into a lot of them, who is doing most
11 of the talking in content area classrooms?
12 Hint. It is not the students. It is, more
13 often we still have the sage on the stage
14 phenomenon occurring way too often.

15 So part of what is being suggested,
16 and what Ed and Lorrie have talked about all
17 really have to do with forms of instruction
18 where the assessment is complementary of that
19 form of instruction. This is critical for
20 this population, because they do not come to
21 school with the vast vocabularies that our
22 native English speakers come with.

23 Because, they may not have

1 opportunities outside of the school, in the
2 school day to engage in more sophisticated
3 uses of language in English. They may be
4 doing it in their primary language. And they
5 have other registers that serves them well in
6 the worlds they walk in. But in terms of
7 academic language to succeed in an academic
8 context, precious little time they have for
9 that.

10 And so we need to maximize their
11 opportunities. We have done studies where we
12 have seen, how much time in the school day in
13 terms of instructional time are kids actually
14 using language? And over and over again, and
15 we do these half day walkthroughs in
16 classrooms with very careful measurements.

17 We find the kids who are talking
18 about 1 to 2 percent of the total time; 1 to 2
19 percent, are kids actually using language.
20 And more often than not, it is simple response
21 tasks. Yes or no, one-word answers, not
22 extended uses of language.

23 So if an English learner performs

1 poorly on an academic content assessments,
2 this is the dilemma. We don't know. Is it
3 due to insufficient language proficiency to
4 demonstrate that content knowledge? Is it due
5 to lack of the content knowledge, meaning,
6 there is an opportunity to learn content that
7 has not occurred? Is it from constructive
8 relevant interference, unnecessarily complex
9 language in the language task.

10 And this extends beyond just the
11 classic large scale assessment. It is a well
12 known fact, and you can watch it with teachers
13 that in grade level tasks or exercises,
14 careful exercises that teachers want groups of
15 students to engage in, in cooperative learning
16 settings, many times if you look at the
17 English learners, it is the language of the
18 instructions that the teacher is giving that
19 is more complex than what the task requires.

20 So think about this. If they can't
21 understand what they are being asked to do
22 because of the language of the instructions
23 are so complex, then they are already off to a

1 bad start. So this kind of issue really can
2 occur in a number of ways.

3 So other sources of advice, or of
4 course, there is cultural distance issues,
5 dialectical variations. Even rater
6 misinterpretations. Again, who is doing the
7 rating of open constructive response
8 assessments and do they understand the issues
9 of second language acquisition and written
10 variation.

11 Now I know it is going to sound
12 like everybody needs to become an applied
13 linguist. I don't want to imply that. What I
14 am suggesting though is that we need this
15 expertise throughout the continuum. But
16 before that, I want to emphasize one thing
17 here. Problems in assessing ELs' academic
18 content, knowledge and skills does not mean
19 that they cannot learn that content.

20 Let me repeat this. Even kids at
21 the lower levels of language proficiency, if
22 they have lower levels of English language
23 does not mean that they can't learn their

1 content. This is not a cognitive issue.

2 And I am saying in English, it also
3 doesn't mean that we don't teach them grade
4 level content until they have sufficient
5 English quote unquote. It means that we need
6 to carefully scaffold that instruction, so
7 that no matter what level of languages you
8 see, kids, even at the periphery have
9 meaningful participation and can engage
10 cognitively with what is being taught.

11 And then the last thing it doesn't
12 mean, is that until they have sufficient
13 English, it is all noise. Actually there is,
14 and this is where computer-adaptive testing is
15 going to come in, we can actually get more
16 information on what kids know if we utilize
17 the kinds of things that Gary was talking
18 about earlier, which is you know, adjusting to
19 difficulty levels to what the kids, where they
20 are responding on the scale.

21 So opportunities for you all. I
22 left myself four minutes to talk about all the
23 rest of this. That was supposed to be ten

1 minutes, but it turned out to be 16.

2 Clearly, what this implies is that
3 we are going to need to revisit the content
4 standards with ELs in mind. And I think that
5 means reenvisioning how the range of content
6 assessments could well, first of all, the
7 standards we will get to. So this is just a
8 quick walk through.

9 Reenvisioning the range of content
10 standards, how they will support teacher
11 practices that strengthen learning. Require
12 students and teachers to engage and reflect in
13 complex tasks and cognitive efforts. Require
14 and promote sophisticated language use. That
15 is among teachers.

16 You can see this in high schools
17 where teachers for affective reasons will
18 start speaking in a different register to
19 connect with their kids. That may be good
20 occasionally. But in fact, they need to be
21 modeling the kinds of language that our kids
22 need to be using in order to succeed in the
23 workplace and succeed in higher education.

1 And then also, we need to
2 reenvision the range of content assessments so
3 that they provide richer clearer signals on
4 what ELs know. And advance new understandings
5 of validity and utility. So what this is
6 meaning, whether it is the common core or your
7 consortium content standards, lets start with
8 reading language arts, even though I was
9 warned not to.

10 Let's just go there. All right.
11 It is going to require a more explicit
12 delineation of academic language forms,
13 functions and communicative competence
14 features. So we were talking about this
15 before.

16 Jim will bring it up about the verb
17 "to know" should be banned from every content
18 standard out there, or "understand and know."

19 They are too vague. They are not specific
20 enough. We need to be speaking about specific
21 functions, hypothesizing, defining,
22 persuading, comparing, contrast, and the kind
23 of things that mediate and also demonstrate

1 content knowledge.

2 And it will involve these other
3 dimensions as well, that are more linguistic
4 and sociolinguistic. And more explicit
5 descriptions of performance expectations in
6 listening and speaking. Meaning the
7 linguistic means of demonstrating content
8 knowledge.

9 We are starting to see this in the
10 K-12 draft common core standards. And it is a
11 good sign. And it needs to be exported, and
12 integrated with other content areas.

13 So where does the ELP stand? Now
14 keep in mind, as these content standards roll
15 out, there are huge implications for the
16 English learners' proficiency standards that
17 are already out there. Those things, many of
18 them have been designed based on current or
19 long past and long overdue for change content
20 standards in the language arts and in other
21 subject matter.

22 We are going to need to make sure
23 there is a stronger alignment between English

1 language proficiency and English language arts
2 standards and other content standards. And
3 also that we have stronger linkage among the
4 assessments.

5 With this federal alignment and
6 linkage, the progress on the ELP assessment
7 may provide greater predictability for ELA
8 performance among our lowest performing ELP
9 level EL students. And this is just a finger
10 pointing in the direction of accountability,
11 where we didn't want to go today.

12 Some promising possibilities for
13 assessing math and science. There are
14 emerging technologies to assess content
15 knowledge of ELs at the lowest ELP levels.
16 One of them is being developed out Wisconsin.

17 I shouldn't advertise. But I just really
18 like it; it is called On Par.

19 And they are doing computer-based
20 dynamic item types that use graphic
21 representation models and sort of a multi-
22 symbiotic approaches. Simulations of content
23 and target stimuli to really get at what kids

1 know who have very little language
2 proficiency. And still can demonstrate the
3 same level of cognitive complexity of math and
4 science knowledge.

5 So it is more expensive, so clearly
6 best used with lowest ELP levels. We are
7 going to need to revisit accommodations.
8 Accommodations research findings have not been
9 encouraging. But I think that is because many
10 of the approaches have been inappropriately,
11 or applied imprecisely.

12 A lot of these came out of the
13 special ed field. They are a much more
14 refined version that will really be effective
15 for English learners. And they are so often
16 aggregated to this monolithic EL level, when
17 in fact, with a level differentiation and
18 diversity in the EL population, we should be
19 really targeted and based on what the
20 characteristics of the particular student are
21 and what they bring.

22 And this method, for example,
23 student profiles being used to assign

1 configurations of accommodations is more
2 promising. And there are efforts underway,
3 that are doing that, and getting good results.

4 Enormous professional development,
5 imperative. We have all talked about this. I
6 think we will just come back to it.
7 Especially around formative assessment
8 development in use.

9 We are going to need to build the
10 literacy, assessment literacy of teachers and
11 administrators, and link to the diagnostic
12 assessments of learning needs, and what the
13 priorities are for instructional decision
14 making and capacity building. So curricular
15 objectives, we will talk about in curriculum
16 embedded assessments for sure.

17 Strengthen the expertise. We are
18 going to need to have experts at every phase
19 of this content development, materials,
20 curriculum framework development, assessment
21 development and professional development.

22 We need educational linguists, and
23 EL assessment and instructional experts to be

1 involved. It doesn't mean that they take
2 over. They have to be a player at the table
3 to help inform our decisions, so that we don't
4 unintentionally miss opportunities and indeed
5 cause harm to these kids.

6 And then finally, well reporting
7 and analysis, I think that we need to be
8 looking at kids' ELP progress by the time they
9 are in programs. Trajectories may vary, based
10 on their level and grade. And we need to look
11 at kids' academic progress and performance
12 based on their ELP level, but also think about
13 timely program.

14 If we have a kid who is at the
15 intermediate level, and he is in the seventh
16 grade, we had better hope that he came in the
17 fifth or sixth grade, because if he came in
18 kindergarten, it is a serious failure of our
19 system to ensure that they have progressed.

20 If we have kids at advanced levels
21 of English language proficiency who are doing
22 poorly in their academic content assessments,
23 that may not at all be a language issue. It

1 may have been an opportunity for access to
2 grade level content while they were learning
3 English issue. And we need to make sure that
4 we address it.

5 And we can through this kind of
6 assessment program development process. The
7 elephant in the room. We were asked not to
8 address it, and for good reason; high stakes
9 accountability. We know that it has perverse
10 incentives. And it affects consequences. No
11 performance without capacity. This is Dick
12 Elmore's phrase. We need to calibrate the
13 demands for performance that we are expecting
14 from teachers with the provision of support to
15 them and for our kids as well.

16 So I think I will stop here, since
17 I have gone over my time. The bottom line,
18 oh. I have got four seconds left. The bottom
19 line. I think we will need to, to not let the
20 perfect be the enemy of the good here. We are
21 going to need with this population
22 particularly to adopt what my colleague and my
23 friend, Luis Hernando Flores says,

1 probabilistic rather than deterministic views
2 of assessing ELs.

3 And by that, I mean, we need to use
4 the assessments carefully, recognize
5 limitations and the ambiguities; strengthen
6 local assessment; don't lower expectations or
7 create separate tracks for these kids; provide
8 useful information about what they know, to
9 improve instructional programs.

10 And don't ignore or downplay the
11 unavoidable problems with this population. It
12 is going to be -- we have to be careful about
13 misleading information. We don't want to
14 unintentionally harm the kids we mean to help.

15 All right. Thank you.

16 MS. WEISS: Thank you.

17 (Applause.)

18 MS. WEISS: Questions? Sure. Go
19 ahead.

20 MS. MELENDEZ: Can you talk a
21 little bit about native language assessments
22 and their role?

23 MR. LINGUANTI: Yes. That is the

1 one thing that I figured that I would leave
2 for tomorrow, so that folks who felt they had
3 heard it all today would want to come back
4 for. It is a huge area. And I think it will
5 be better discussed tomorrow.

6 But in a few phrases, the question
7 is, what is the proper -- reinterpret your
8 question, and stop me if I am wrong. What is
9 the proper role for assessments of content in
10 the students' native language? When is it
11 appropriate? What can we learn from it? Is
12 it just to be used in summative assessment?

13 And I think the answer is very
14 complex, because kids come to us, as I tried
15 to indicate before. Some ELs come to us, and
16 they are actually English dominant. They may
17 have some oral proficiency in their primary
18 language. But they may be much more
19 preferring to use English or stronger in
20 English.

21 Some kids are actually being taught
22 in the primary language, which is a very
23 important point. But they may still be

1 accessing certain other kinds of knowledge in
2 the second language as well. So I don't think
3 there is a very simple cut and dried answer
4 about if -- even to this level, if instructed
5 in the primary language test and the primary
6 language, I think it is more complex than
7 that.

8 But for sure, when kids are being
9 instructed in their primary language, in dual
10 language settings, we should certainly be
11 building on assessments to tap their
12 knowledge, to understand what they are
13 learning in that language, if that is the
14 instruction. But also don't forget they may
15 actually have knowledge that can be displayed
16 and demonstrated in English. And they may be
17 working back and forth between those.

18 The other role for primary language
19 assessments. Certainly, at high school level,
20 if you have kids coming in, more recent
21 immigrants, we need to get a better idea of
22 what these kids are coming with. And primary
23 language assessments can certainly help us get

1 a sense of, in terms of their knowledge base
2 and their cognitive skills and thinking, where
3 these kids are.

4 And if it is a diagnosis for
5 special ed, I have kids that appear to be
6 presenting learning difficulties, primary
7 language assessments are very important. So
8 there are key roles in here. Informative
9 assessment, I would say, it really will be
10 probably more driven by what the language of
11 instruction is in the classroom.

12 But we need to not fall into the
13 binary approach of, well, it is only in
14 primary language or it is only in English.
15 And there needs to be this sort of quick
16 transition dividing line.

17 I think we are going to need to be
18 more sophisticated than that. If for nothing
19 else than because our instructional programs,
20 dual immersion programs are among the fastest
21 growing of bilingual programs where half the
22 student population are native English
23 speakers, half the population are native

1 speakers of the minority language. And they
2 each learn each others, they each learn
3 academic content using each others languages.

4 And what the goal is, bilingualism and
5 biliteracy for both sets of students.

6 And clearly there, you need to
7 think about using Spanish language assessments
8 for language minority kids of the Spanish
9 language, that is, native English speakers.
10 So there are -- I think our instructional
11 programs are out ahead of our assessment
12 capacity, certainly in large scale right now.

13 We need to think about where that fits in
14 this.

15 MR. SMITH: Follow-up on that last
16 comment you just made. It occurred to me
17 while you were talking about the lack of
18 language in the classrooms. This is not just
19 a problem of English language learners.

20 MR. LINQUANTI: No indeed.

21 MR. SMITH: This is a problem that
22 pervades classrooms and schools of high
23 poverty. It is unbelievable. You get that

1 same one and two percent when people do
2 evaluations of this. And you don't have --
3 and it is the nature of the language, not to
4 make --

5 MR. LINGUANTI: The quality of the
6 language.

7 MR. SMITH: But it is the quality
8 and nature of the language. And a lot of this
9 is wrapped into academic language as well. So
10 the academic language issue is not just an
11 issue of English language learners.

12 It is a fundamental issue of an
13 awful lot of, are our childrens learning. So
14 that raises the question in my mind, do we try
15 to diagnose? Do we use assessments to try to
16 better understand the quality of the academic
17 language that the kids have as they come in?

18 So this is not just measuring what
19 you have learned in second grade. It is
20 measuring whether you know what greater than
21 is, maybe for fourth grade or fifth grade or
22 whatever.

23 Uri Triesman has done a bunch of

1 studies on his algebra work, and he finds that
2 half the errors that are made in algebra, in
3 eight grade on his students are made because
4 they didn't understand the problem. They
5 understood the algebra, but they couldn't put
6 the numbers and the formulas together in the
7 right way because they didn't understand the
8 language.

9 MR. LINQUANTI: Yes.

10 MR. SMITH: This may be something
11 you want to talk about. This is a different
12 kind of assessment. But it is clearly part,
13 potentially, of an assessment system of all
14 kids in the general sense.

15 MR. LINQUANTI: And I would just
16 respond briefly to that. That there is, as
17 way of clarification, that many folks who are
18 working with non-English learners look at some
19 of the academic language tasks that are being
20 asked of English learners as part of their
21 second language proficiency building and say,
22 hey, all our kids need that. That would be
23 good for our kids too.

1 So I think what we need to do,
2 because there is a risk there, in that, it is
3 saying, oh then these kids linguistic needs
4 will be met by everyone. We have to be very
5 careful. Each teacher has an obligation to
6 instruct and ensure that students learn the
7 language of their discipline, of the teachers'
8 discipline.

9 If they love their discipline, they
10 should be really focusing on the language,
11 informs, discourse patterns in which that
12 discipline takes place. But there are unique
13 needs of second language learners as well.

14 And then there are needs of
15 speakers of other varieties of English, you
16 know where there is African-American
17 vernacular English or Chicano English or those
18 things. Or just kids that don't have access
19 to this formal register of English where that
20 needs to be practiced.

21 So I think they are interrelated.
22 But there are still going to be things that
23 are very specific to non-native speakers of

1 English that need to be really focused on as
2 well.

3 MS. WEISS: Great. Thanks. I am
4 going to actually move us on to Jim.

5 MR. PELLEGRINO: Okay. Well, I am
6 the last in the line-up. It is about the time
7 in the afternoon when people are getting
8 drowsy. So but I also get the benefit of
9 being the last one. So I get the recency
10 effect.

11 A disclaimer. I am not a
12 psychometrician; I am a cognitive
13 instructional psychologist. I just hang out a
14 lot with psychometricians for some strange
15 reason.

16 I am really delighted to have a
17 chance to share with you some thoughts on the
18 whole issue of assessment systems and how we
19 might think about them in the context of the
20 Race to the Top. There is a series of issues
21 and topics that I just want to run through
22 that run from some general issues about design
23 and thinking about the design and the

1 components of the system to some issues that
2 have to do perhaps closer to the way in which
3 the RFP is written or some of the things that
4 are built into the RFP.

5 So I will try and chug through each
6 of these eight, and see if I can make some
7 reasonably coherent points. You will see a
8 lot of commonality, and hopefully will have,
9 with remarks made by my colleagues earlier
10 today. No great surprise. Like I said, I
11 hang out with these folks a lot. So some of
12 it rubs off, for good or for bad.

13 So the first thing I want to talk
14 about is what I think are, what should be the
15 elements in the comprehensive system. What
16 are we trying to design? And I think that the
17 important idea is that the desired end product
18 that we want is a multilevel system.

19 And in that system, each of the
20 levels has to fulfill a particular set of
21 functions. And it recognizes that there are a
22 clear set of intended users of the
23 information.

1 So one of the mistakes we make is
2 somehow thinking that everybody is going to
3 use the information at every level of the
4 system. There are teachers that need one
5 level of information. Superintendents then
6 might need another. And policy makers then
7 might need another.

8 So the system needs to be designed
9 to meet the information needs of those
10 individuals at those levels. And the
11 assessment tools are designed so that they
12 serve that intended purpose. We continue to
13 sort of operate within an assumption that we
14 can accomplish multiple purposes with the same
15 assessment, rather than realizing that we have
16 to design an assessment to fulfill a function
17 like formative or summative or accountability
18 purposes.

19 And so the design needs to be
20 optimized for the particular function that you
21 are intending to serve. And the other part is
22 that these levels, then, while they need to be
23 there, they need to be articulated and

1 conceptually coherent.

2 That is, they have to have the same
3 underlying concept of what the targets are, of
4 learning at a given grade level, and what the
5 evidence of attainment should be. But the
6 difference is, they provide information at the
7 grain size, and on the time scale that is
8 appropriate for the intended user of that
9 information to translate that information into
10 action.

11 Those time scales are going to
12 vary, and very short time scales for classroom
13 teachers to much longer time scales at the
14 level of state and federal policy makers.
15 Here is an example that comes from the former
16 NSF center on assessment and evaluation of
17 student learning. It just gives an idea of
18 what a multilevel system might look like, with
19 three levels.

20 And the idea, it is an integrated
21 system, in which there is coordination across
22 the levels that is unified by common learning
23 goals. And it is synchronized by unifying

1 progress variables. Earlier today, I think
2 Lorrie used the language that we had in the
3 state science assessment report of horizontal,
4 vertical and developmental coherence.

5 And so these are things that the
6 system should be designed and aspired to. The
7 system should be designed to track progress
8 over time. And by tracking progress, we
9 probably will need to do it at the individual
10 student level, as well as the aggregate group
11 level. Doing it at one level may be different
12 than doing it at another level.

13 For example, we track progress at
14 the aggregate group level, using things like
15 NAEP. NAEP does not give us individual
16 scores. So we can use multiple models. The
17 idea is that the system that we build has to
18 have tasks and tools and technologies that are
19 appropriate to the desired inferences about
20 student achievement that we need to make to
21 fit a particular purpose.

22 We can't force everything into a
23 fixed testing or task model. There isn't a

1 sort of magic bullet that says we should use
2 this task for this purpose. We should use a
3 range of things, performance tasks,
4 portfolios, projects, fixed and open response
5 tasks as needed to fulfill the goals that we
6 have in mind, or the purpose that we have to
7 have at each of these levels.

8 The second point I want to make is
9 one you have heard. Integrate assessment with
10 curriculum instruction. Assessment has to be
11 part of a coordinated system that includes
12 explicit assumptions about curriculum and
13 instruction.

14 And in fact, all three of these
15 elements have to work in concert and have to
16 be designed as compliments of each other,
17 rather than having a tension with each other.

18 I just put up there, my way of thinking about
19 what is curriculum and what is instruction and
20 what is assessment.

21 Because I realize that when I
22 talked about curriculum, I may use the term
23 differently than others. The point is, is

1 that these three elements need to be thought
2 through in terms of how they relate to each
3 other and in essence, how they fit together.
4 Lloyd said this earlier. There is your dog,
5 Lloyd, and we don't want assessment to wag the
6 educational dog.

7 Instead, what we want is something
8 that in fact has alignment among curriculum
9 and instruction and assessment. And notice
10 that in the middle of that, I put, I don't put
11 standards there. I put a theory of learning
12 and knowing. In fact, because I am going to
13 say a little bit more about standards.

14 What should drive or design of the
15 three things is what we know about the nature
16 of what it means to understand and know
17 something in a particular content area. And I
18 will say more about know and understand in a
19 minute. And what we know about that
20 empirically, in terms of how this is acquired
21 by students over time under conditions of
22 instruction.

23 Now, common core standards are a

1 start, but they are not enough. Standards are
2 problematic, because getting people to agree
3 on standards means that they are probably
4 ambiguous.

5 And they are oftentimes stated in
6 terms of two verbs, "know" and "understand."
7 And the reason that people can't agree on
8 standards is because your knowing is different
9 than my knowing; your understanding is
10 different than my understanding.

11 And so we think we each understand
12 each other. And we never have to get down to
13 brass tacks. One of the things that I have
14 come to understand much more deeply is that
15 standards need to be translated using a
16 processes like backwards design and evidence-
17 centered design. In a way that actually moves
18 us forward to address this issue of curriculum
19 instruction and assessment.

20 And I am going to use the terms
21 that we have used from evidence-centered
22 design in which you unpack the standard, to be
23 clear about the claims you want to make about

1 students as a result of particular
2 instructional experiences -- the evidence that
3 you would use to back up those claims.

4 And then you talk about the tasks
5 or situations that would give you that
6 evidence. We tend to jump from standards to
7 assessments rather than unpacking the
8 standards and saying, what does it -- when we
9 say the students should know proportional
10 reasoning or understand proportional
11 reasoning.

12 We have to unpack that, in terms of
13 the specific things that students are able to
14 do. There are good examples of actually how
15 to engage in this kind of a process at scale.

16 And I am going to mention one that I have
17 been involved with for the last three years,
18 which is the design of AP science courses and
19 exams as well as the College Boards recent
20 effort to put out standards for college
21 success derived from that.

22 And in essence, what we did in that
23 work is to try to lay out a domain model. A

1 domain model for physics and chemistry and
2 biology and environmental science that is very
3 explicit about the claims you want to make
4 about students at the end of a course.

5 The supporting evidence and the
6 different levels of achievement that you would
7 expect to see before you go and build the
8 test. And in fact, it is a way to build an
9 assessment framework as well as a curriculum
10 framework and to figure out essentially, okay,
11 what does the test need to look like, rather
12 than working the other way around, which has
13 always been with the AP, which is we build the
14 test. And then teachers have to figure out,
15 okay, what do I need to teach? Because that
16 is the test.

17 The third thing I want to touch on
18 is essentially the idea of considering
19 alternative design options for K-8 and high
20 school. Basically, K-8 needs to be designed
21 to prepare students for success in high
22 school. Not just to get them into high
23 school.

1 And actually, we have evidence now
2 that many students, even though they exit
3 eighth grade performing above norms on
4 standardized tests are not prepared to succeed
5 in high school. And in fact, have a very low
6 probability of success on the ACT at a level
7 that is going to get them into a good college.

8 And there has been some nice analysis of
9 that, even in Chicago.

10 So what we need to have at the K-8
11 level is a clear, coherent and cumulative
12 model that links curriculum, instruction and
13 assessment in each major subject matter area
14 and it incorporates learning progressions and
15 benchmark targets at critical age and grade
16 levels. Our assessment practices need to
17 support the desired intellectual performance
18 we want. That is, it shouldn't undermine it.

19 It should be compatible with it, and actually
20 promote it.

21 So we should choose tasks and
22 performances that are matched to the desired
23 intellectual performance demands that we want,

1 rather than what is easy to do. And I think
2 Ed or someone may have alluded to this
3 earlier. We need to consider multiple options
4 to maximize the information yield for
5 different intended levels of use.

6 A mixture of census testing and
7 matrix sampling may be optimal, particularly
8 if we want to get at some of the kinds of more
9 complex forms of reasoning, where we cannot
10 possibly test all kids on all items that we
11 might want. So we might want to have a
12 mixture that uses some mixing of census and
13 matrix sampling.

14 We also should consider some
15 options that include periodic versus yearly
16 assessments in core subject matter areas. At
17 the high school level, I think again, high
18 school should be designed to prepare students
19 for college or work. But the same model of
20 testing that we carry over from K-8 may be
21 completely inappropriate and suboptimal at the
22 high school level.

23 Not all students take the same

1 courses at the same time. So we need to
2 consider alternatives, such as assessment that
3 is more closely tied to curriculum and
4 instruction frameworks in core subject matter
5 areas. And use exams that are tied to
6 courses.

7 We also need to consider options
8 for certification of college readiness and
9 more rapid transitions to college or into
10 vocational programs in lieu of our current
11 model, which is essentially tie them in seats,
12 and numbers, of course, is then taken.

13 So in essence we need to have
14 flexibility rather than rigidity at the high
15 school level, in terms of how we think about
16 assessment and how we think about moving
17 students from high school on to college and on
18 to work. We heard a little about technology
19 today. I want to make a very strong push on
20 capitalizing on the multiple affordances of
21 technology.

22 What technology can do for us, in
23 terms of assessment is a very wide range of

1 things. And we no longer need to be wedded to
2 the paper and pencil technology that we have
3 had for the last almost 100 years. And in
4 fact, one of the real problems is that all we
5 have done with technology for the most part is
6 to move our paper and pencil tests onto
7 computers and have done very little in the way
8 of anything innovative with some exceptions
9 like computer-adaptive testing.

10 So there is a variety of things
11 that technology can allow us to do, which will
12 allow us to actually ask and present more
13 challenging, interesting problems; collect
14 much more interesting and rich forms of data,
15 and to do various kinds of complex scoring.
16 And there are functioning models of technology
17 for things like NAEP or uses of technology for
18 large scale programs like NAEP and PISA.

19 It is interesting that the places
20 where there is innovation, are places where we
21 haven't had high stakes testing. At least,
22 not high stakes at the individual student
23 level.

1 The other thing I want to mention
2 is something that has been said already, is
3 that technology allows us to adapt the
4 assessments to the needs of different learners
5 and student populations. It may actually make
6 or enable the use of universal design
7 principals much more so than we can do now.

8 And it can implement better
9 adaptive assessments and computer-adaptive
10 testing methods. If you haven't seen it, this
11 is an example. This is a still of a kind of
12 item that was a prototype for NAEP science,
13 that has students engaging in an interesting
14 inquiry problem.

15 The question is up there at the
16 top. How do different payload masses affect
17 the altitude of a helium balloon? They can
18 manipulate things, control variables. Look at
19 results. Plot data. And they can construct
20 an argument.

21 Now I am not saying this is a
22 perfect kind of item. But there, I want to
23 use this to illustrate the point that we can

1 put situations in front of students that are
2 much richer in terms of the kinds of tasks we
3 ask them to engage in. Robert mentioned
4 something about the work in Minnesota. This
5 is actually, I think, one of their items.

6 Which again, is an item that allows
7 students to look at a complex ecosystem. And
8 it is being used with English language
9 learners and other students because it has the
10 capacity to adapt to their needs. And at the
11 same time, present them with an interesting
12 and challenging problem.

13 Some of the other affordances of
14 technology; it makes possible to embed
15 assessments in learning environments. It
16 actually allows us to implement some things at
17 the level of diagnostic, informative
18 assessment that are otherwise difficult, if
19 not impossible to do easily in regular
20 classrooms.

21 And there already exists a number
22 of workable systems and models for use in
23 schools, and areas of literacy and mathematics

1 and science. I will mention three in a
2 minute.

3 The other thing I want people to
4 think about, because technology is not one
5 thing. It is many things. And it can fulfill
6 multiple functions in an assessment system.
7 And we need to think about all of the things
8 it can do to support the design implementation
9 and use of a multilevel next generation
10 assessment system.

11 For those of you who are
12 interested, these are some examples of
13 formative assessment systems that are used
14 right now. One of the things that is
15 interesting about them is that they all
16 incorporate some notions about the nature of
17 student learning, and misconceptions. And
18 they use it as a diagnostic basis with the
19 role of supporting teachers in the classroom.

20 And in terms of those different
21 functions that I was talking about, these are
22 just the way to sort of think about the
23 multiple functions of technology in an

1 assessment system. In terms of the design and
2 the collection of information, the reporting
3 and analysis of information as well as the use
4 of that information to support instruction.

5 So there is a wide range of things
6 that technology can do. I am getting there.
7 I have some time left still. We haven't said
8 this today. I want to make a pitch for making
9 science an equal partner in this. And my
10 argument has multiple components.

11 First of all, science at the K-8
12 level of the United States is a mess. It is
13 in absolute and desperate need of improvement.

14 We have no coherent model for what is to be
15 taught and learned in K-8 science. And
16 international assessments show that the
17 further we go out, the worse it gets for us.
18 And it is one of our weakest areas.

19 Leaving science out of the system
20 contributes to a second tier status. It
21 allows it to be treated as a periodic rather
22 than ongoing and critical instructional issue.

23 Pity the poor fourth grade teacher

1 who happens to be in a state where this is the
2 grade level at which science is tested under
3 No Child Left Behind. How many teachers don't
4 want to teach fourth grade, because that is
5 the grade at which science is tested.

6 Now the other thing is to think
7 about science, not just as science, but the
8 fact that it supports the development of
9 verbal and quantitative literacies. Much of
10 science is really about the process of
11 reasoning and arguing from evidence. And
12 doing so requires both qualitative and
13 quantitative reasoning and communication
14 skills.

15 It is not just about little bits
16 and pieces of facts that have to get
17 regurgitated. It is about reasoning from
18 evidence. And in a 21st century global
19 environment, how can you ignore the scientific
20 and technical literacy of your students.

21 And the last point I want to make,
22 the pitch I want to make for science is that
23 science achievement at the high school level

1 suffers from the poor preparation of students
2 at K-8. Those who teach at the high school
3 level essentially assume little in the way of
4 carryover or cumulative knowledge, coming into
5 high school. So it significantly affects what
6 we can do at high school.

7 Not to mention the fact that we
8 seriously need to rethink what science needs
9 to be taught at high school in the context of
10 contemporary changes in the nature of science.

11 We are still with the same model. Biology,
12 chemistry, maybe physics. A little earth
13 science. Maybe some environmental science.
14 But it is not a very coherent model.

15 And most of what is taught is
16 actually not at the cutting edge of science in
17 terms of molecular biology, nanotechnology, et
18 cetera; phase into components and build system
19 capacity; consider the design and
20 implementation as a seven- to ten-year
21 comprehensive effort and long-term investment.

22 Basically, the argument I want to
23 make here is, you can't do, go to all these

1 components together in a short time frame.
2 The capacity to do this, not at the scale that
3 is needed. It is just not there to do it all.

4 So instead, identify key leverage
5 points that can be phased in over different
6 time horizons. If we rush states to do it
7 all, the result may be no significant
8 improvement. The unintended consequence I
9 would like to see us avoid is what has
10 happened with No Child Left Behind, which is
11 essentially we set assessment practice back
12 ten years. We can't afford to set it back
13 another ten years. And consistent with what
14 others have said, part of this is about
15 building capacity. Human capacity and system
16 capacity. And this includes teachers,
17 administrators and others.

18 Basically, what we want is to shift
19 from an audit mode for an assessment to an
20 assistance mode. That is the considerable
21 mind shift for most educators and most policy
22 makers. And we need to have systems that can
23 support that. I am a little over my time.

1 But I will quickly finish up here.

2 I think that states need to be
3 asked to lay out a logic model for system
4 design, implementation and evaluation. A
5 logic model forces people to attend to the
6 existing conditions, the resources, inputs,
7 outputs and outcomes.

8 And not only do you focus on the
9 elements of the system, but their logical and
10 causal relationships. It enables you to
11 monitor the building of this system, its
12 enactment. And it gives you a way to think
13 about evaluation and adjustment, because you
14 are putting in place a complex system.

15 I will skip this. But this sort of
16 illustrates the components. I am not saying
17 this is the model. But this has the
18 components that you need to think about in
19 terms of what is connected to what. What it
20 does is, it forces you to specify your theory
21 of action. Okay.

22 And your theory of action is really
23 important because, in essence, it is the

1 theory of action that forces you to make
2 explicit about how particular components are
3 supposed to work, who is supposed to be
4 impacted, what is the expected consequence
5 supposed to be, and then why. And then if it
6 doesn't work that way, in essence you can test
7 it, as long as you also lay out different
8 competing theories of action and what that
9 would imply.

10 Lastly, policy articulation between
11 Race to the Top and ESCA. The bottom line
12 here is, that there is a lot of policy
13 implications and policy choices that will have
14 to be made at the federal, state and local
15 level with respect to regulations and funding.

16 You need to be clear what the
17 policy issues are, and what the unintended as
18 well as intended consequences are of some of
19 those things. I also think you need to be
20 clear about listening to unbiased groups like
21 the NRC's Board on Testing and Assessment when
22 it offers serious cautions about things like
23 value-added methodologies.

1 And you have got to think about how
2 is all of this going to work with ESCA,
3 because ESCA is still the 800-pound gorilla
4 that is out there, and it is going to continue
5 to be out there. So I will quit there.

6 (Applause.)

7 MS. WEISS: Great. Any quick
8 clarifying questions?

9 MS. WHALEN: I had a quick one. At
10 one point in your presentation, you identified
11 that we may think about associating critical
12 ages earn critical weights? In your system or
13 in your assessment system, what would be the
14 optimal grades to be tested or what are those
15 critical bench, gateway places that you would
16 want us to focus in on?

17 MR. PELLEGRINO: Well, in some
18 ways, we do that know, because we have a thing
19 called NAEP. So we have to think about first
20 of all, where NAEP fits in this entire system
21 as well. And is it going to be this sort of
22 outside the system.

23 The answer to your question is, in

1 part, I think it depends also on as we lay out
2 for these particular curricular areas, what do
3 we see as particular key benchmark points? So
4 four and eight may not be sufficient.

5 It might be that we want to test
6 you know, three or four times across the K-8
7 span for particular things. For example, in
8 reading, we might want to test somewhere
9 around the second grade. And then we might
10 want to test again at the fourth grade and the
11 sixth grade and eight grade, if we were really
12 articulate about the transitions, in terms of
13 learning to read, versus the reading to learn
14 issues and the comprehension issues.

15 So I think it is going to vary
16 depending upon the content areas. I think if
17 you ask the disciplinary experts as well, they
18 might also have a different model for what to
19 do periodically for particular aspects of the
20 mathematics curriculum in terms of things like
21 number and algebraic, you know, prerequisites
22 for algebraic understanding. I can't answer
23 that in the abstract.

1 But I think it should be easy to
2 identify those things, and to do those well,
3 perhaps not to try to do it at the individual
4 census testing level, but more at the
5 aggregate level for monitoring, because you
6 are going to do a lot of individual testing
7 anyway at the level close to instruction.

8 MS. WEISS: Do we have any other
9 quick clarifying questions? Because I want to
10 move into our roundtable discussion pretty
11 soon.

12 So yes, go ahead Mike.

13 MR. SMITH: Yes. Were you
14 suggesting that the eighth grade, at the high
15 school level, are you suggesting the standard
16 course exams as your tool?

17 MR. PELLEGRINO: Yes. As one
18 option, rather than just trying to come up
19 with a generic math test or -- because in
20 essence, it is not clear what those things,
21 how they fit with anything in the curriculum.

22 MR. SMITH: Right. And you could
23 do that in the so-called academic track and in

1 vocational courses as well.

2 MR. PELLEGRINO: Yes.

3 MS. WEISS: Yes. Well, this is
4 just a tiny point. But whenever there is this
5 much of the Department represented, I would
6 like to get this thought put in the back of
7 your minds.

8 In the ESEA context, back to
9 Robert's comments, it is analogous to the
10 reform that has happened now with the
11 Governors; to redefine how they think about
12 dropouts. That you get inaccurate information
13 if you do an annual cut. You must track it
14 longitudinally, or you don't understand it.

15 Similarly, if you want to study the
16 progress of second language learners, you must
17 track them across that artificial dichotomous
18 category. And that is not necessarily about
19 this. We could talk about how learning
20 progressions relate to this. But that is a
21 very critical thing to be resolved in other
22 arenas.

23 MS. WEISS: Great, I think, let's

1 use this as the transition point, sort of
2 moving to our roundtable discussion.

3 And let me just pick up where we
4 were going, because Lorrie I think said that
5 her premise was that to get it right we needed
6 to go all the way to the ground. We needed to
7 deal with standards, curriculum and
8 assessment, or curriculum, instruction and
9 assessment, as Jim put it. And that would
10 make it impossible to take on everything.

11 So we need to take on small pieces.

12 Jim for perhaps different reasons said we
13 can't do it all, or we will get it all wrong,
14 and so we need to pick small pieces. So it
15 sounded like you might have different visions
16 of what those small pieces look like and
17 different reasons for why.

18 But can you guys tell us more about
19 what you think, how much is the right amount
20 to take on, and why it is that that right
21 amount will still help us have a coherent and
22 thoughtful and comprehensive assessment
23 system, if we do it in this way. So, Jim, do

1 you want to just sort of --

2 MR. PELLEGRINO: Well, I am not
3 sure that my right amount is the same as
4 Lorrie's right amount.

5 MS. WEISS: Right. Me neither.

6 MR. PELLEGRINO: I think that when
7 I was -- one of the things that I was trying
8 to say is that to build all the components of
9 the system, and to actually build the quality
10 assessments and everything that needs to be
11 articulated is an incremental task. That you
12 can't put in place all.

13 You can't be assured that four
14 years out, you are going to have all of those
15 things. And they are all going to perfectly
16 well articulate. So that is one of the things
17 that I was sort of saying is figure out what
18 we can build reasonably within a three- to
19 five-year time frame. Which components then
20 need to be added on to that, because add
21 another set -- there is things that we don't
22 know in terms of even some of the measurement
23 issues.

1 And although there is a lot that we
2 can do to experiment to get us to that point,
3 including technology. But the thing that I
4 would think about is, I believe that if you
5 take a segment of the curriculum.

6 That is, for example, I think that
7 you could reasonably think about trying to do
8 something K-8 in mathematics. And think about
9 that as a sort of reasonable chunk.
10 Particularly since mathematics, a lot of work
11 has already been done.

12 I would contrast mathematics, K-8
13 mathematics with K-8 science. K-8 science is
14 nowhere near where K-8 mathematics is. So
15 there are parts of the curriculum at different
16 at grade levels where I think there is a real
17 honest to goodness shot of building this
18 coherent model, and putting it in place,
19 curriculum instruction and assessment, in a
20 reasonable time frame. That to me, is a
21 reasonable chunk.

22 MR. SMITH: Let me follow up on
23 that just a little bit. Imagine that we are

1 with the consortium now. We are with the
2 consortium and we each represent a state. And
3 we say, look. We want the following.

4 We want a really good job done on
5 K-8 mathematics. We want a really great job
6 done on English language learners. And on ELL
7 and English language in general. And we want
8 something in science.

9 And we divided it up. And you guys
10 were in charge of math and you guys are in
11 charge of language learning. And you are in
12 charge of science. And why would -- why do we
13 need separate consortia that are doing that,
14 when one consortium if it is large enough
15 could do it, itself.

16 MR. PELLEGRINO: It would have to
17 be a hell of a consortium.

18 MR. SMITH: You are thinking of --
19 I mean, I don't see the distinction. Should
20 you divide up the tests? It could be divided
21 up either randomly among the various clusters
22 of states, or it could be divided up by a much
23 larger group of states, and done in a much

1 more coherent fashion.

2 MS. WEISS: And before we get into
3 the question of how do you organize consortia,
4 I think the question that Mike is trying to
5 ask is, are there chunks that would lend
6 themselves to coherence or not, that we need
7 to be aware of. Are there ways that we need
8 to think about, if we allow people to chunk it
9 up, are there ways that we should allow it or
10 shouldn't allow it in order to make sure that
11 we end up with what we are trying to get here.

12 MR. SMITH: Yes. It is partly
13 that, Joanne. But it is also partly that this
14 distinction between somehow separate
15 consortia --

16 MS. WEISS: Right. But I just want
17 to get into the consortium questions, where
18 there is a separate question and stick for a
19 minute with what is --

20 MS. SHEPARD: Well, first of all,
21 Jim and I do not agree at all about
22 instruction. Instruction has to be in there
23 also. So curriculum, assessment, and

1 instruction, absolutely, we agree. And I
2 think we would chunk things similarly.

3 And I would refer you to the same
4 website I gave you before for the math and
5 science white paper that talks about the need
6 in K-8 science curriculum development along
7 with supports for instruction and a coherently
8 aligned assessment.

9 MS. WEISS: But Lorrie, your chunk
10 sounded way smaller than what Jim was talking
11 about.

12 (Simultaneous discussion.)

13 MS. SHEPARD: Well, I think I
14 had -- well, I am not in favor of high stakes
15 summative assessments at grade one and two,
16 which is the only limitation. My math example
17 was four through eight. So we are very
18 similar in how we would chunk the strands.

19 And depending on whether you were
20 making an accountability test of it, I would
21 say yes. K-8 science curriculum development
22 makes sense. Writing could be even across
23 more grades. You could have smaller consortia

1 do courses in high school, including things
2 like high school certificates, because those
3 are more manageable tasks. So I don't think
4 we are disagreeing.

5 Now to Mike's question, should it
6 be one consortium, and they divide up, that is
7 an open question. I do not claim to know the
8 answer to this. But watching the mess that is
9 made by politics, I fear what will happen if
10 people take on more of a handful than they can
11 mediate through.

12 Then they can commit from the
13 beginning that they will do it in a
14 fundamentally different way, because
15 otherwise, we are going to end up with what we
16 have now.

17 MS. WEISS: Yes. Go ahead, Ed.

18 MR. HAERTEL: I think an analogy
19 that comes to mind is some of Bridget Baird's
20 work on collaborative group work, where the
21 children need to simultaneously negotiate a
22 social structure and a task structure. They
23 need not only to understand the stated plan,

1 with respect to the problem, they also need to
2 understand the state of play as to when that
3 idea is going to be taken up by other
4 participants in the groups, and how to get
5 their ideas there and so forth.

6 The problem of building a
7 consortium is a social problem. It has a
8 political dimension of course.

9 We don't really know how states
10 ought to work together to bring in all the
11 right constituents to get agreement around the
12 table. Your model, Mike, seems to be
13 predicated on the idea that states would be
14 willing to enter into subsets within the
15 consortium. Be willing to enter into an
16 agreement that if you do this well, and let me
17 do that well, I am willing to buy your version
18 of that, and you buy my version of this. That
19 may happen.

20 But oftentimes, you find that you
21 can get consensus around the table among a
22 group. But then when you take out the wide
23 world, the people who aren't at the table, who

1 weren't privy to those conversations shoot it
2 down.

3 So it seems, my intuition, and lord
4 knows, I am a psychometrician so I have no
5 social allusions whatsoever. But the way I
6 would approach it would be to get everybody to
7 agree on something, learn how to work
8 together, and then to build up from that.

9 And that also might -- another
10 piece of that might be to build in an explicit
11 proposal, request in the RFP that people give
12 you an explicit proposal for the staged
13 transition from what they have now to what we
14 are going to have when we are done: what is
15 going to be done first, what is going to be
16 done second, and so on.

17 And that way, you might actually
18 leave something enduring after the funds are
19 expended at the end of four years that will
20 allow the states to move forward. A continual
21 process, using mechanisms and social
22 structures and decision processes and flow
23 charts that they put together.

1 MR. SMITH: Yes. I think both
2 ideas are great. The latter one is even
3 greater. I mean, there has to be some sort of
4 logic model to pick up to people's
5 perspectives on this, of what happens in that
6 transition period. How do we get from here to
7 there.

8 I am still not convinced that you
9 couldn't do this as a larger consortium and
10 you know, taking the problem apart, the
11 consortium itself could come to some decisions
12 about these things, and then allocate out
13 their work in a way that, a priori, that
14 everybody agrees in some sense or way. I
15 mean, that is of course what nations try to
16 do, nations and children, and so on, and maybe
17 states and maybe none of them could do it.

18 But I think it would be worth a
19 try, because the other part of it is pretty
20 distant. It happening over here. And they
21 are much less likely to feel an ownership for
22 it than they would otherwise.

23 MS. WEISS: Yes. Lloyd.

1 MR. BOND: I am sitting here
2 thinking out loud. We need a mechanism for
3 having these consortia talk to each other,
4 because otherwise, they may go off into -- I
5 mean, we need some kind of perhaps some kind
6 of executive panel to have them come together
7 occasionally, in a way, to make sure that
8 they're, you know, on the same page here.

9 MR. LINQUANTI: Some facilitative
10 body, or --

11 MR. BOND: Yes.

12 MR. PHILLIPS: May I respond. Yes.
13 I would think the best of all worlds would be
14 one consortium. That would be great. That is
15 not going to happen, probably. But hopefully,
16 there would be some large consortia. And I
17 think this would work.

18 All of these ideas would work
19 better with larger consortia, I believe.
20 However, I also think there needs to be
21 hopefully something in the reauthorization
22 that addresses this issue, this
23 organizational, structural issue of how to

1 manage all this, and move it along. The
2 states themselves could create something like
3 an OECD, for example, would be a good
4 organizational structure where they all pitch
5 in.

6 And there is an executive group.
7 And it represents everybody. And they manage
8 this. The federal government could impose
9 something like a NAGB body or something that
10 is independent of the government but still
11 oversees it in accordance. So there is a
12 number of organizational structures.

13 Without that, I can see a lot of
14 chaos developing along the consortia. So that
15 has to be addressed immediately, and hopefully
16 within the reauthorization.

17 MS. WEISS: Yes. I think you are
18 right. I think we are thinking about that.

19 Yes, Jim?

20 MR. PELLEGRINO: Just one thing
21 that speaks to Mike's issue. There is also a
22 danger, by having groups sort of work
23 separately on and not communicating with each

1 other on math and science and literacy, which
2 is you essentially build something that is
3 sublime in your area, and you ignore
4 everything else.

5 So in essence, you build a model
6 that is nothing, there is room for nothing but
7 math, and you avoid thinking about the serious
8 cross-talk that is possible, in terms of how
9 these things can enforce, reinforce each
10 other, particularly as you move up through
11 grade levels with respect to forms of
12 literacy.

13 MS. WEISS: So let me just broaden
14 this then, and take it into the consortium
15 questions that we had. Which is, if you were
16 states trying to put together a consortium,
17 what would a good consortium look like? What
18 would the features be that you would look for?

19 And if you were the Department trying to
20 write a notice that said here is what you get
21 points for.

22 It is a competition. Right, you
23 get points for doing consortia that have these

1 features or these qualities, because we think
2 this is what will lead to something that is at
3 least enduring enough to take you through the
4 four-year development cycle of this, and
5 transition us to the OECD of the states.

6 MR. BOND: You are asking a
7 psychometrician this?

8 (General laughter.)

9 MS. WEISS: So let me ask the
10 others then to answer.

11 MR. LINQUANTI: Sure. I am happy
12 to take a crack. I mean, one of the pieces,
13 and it occurred to me during lunch was, what
14 kind of facilitative processes and supports in
15 terms of kind of a technical assistance to
16 ensure that just at an organizational level,
17 it can be very difficult to have a consortium
18 start up, and get functioning. So you know,
19 who would play that role?

20 Maybe it is not any one member of
21 the states. Maybe it is a contributing body.

22 But maybe it is an outside facilitative,
23 technical assistance provider. If they have

1 the capacity to do it. Could help that.

2 Because that tends to allow states
3 to be participants rather than having to take
4 on a lot of the management functions. Again,
5 now that grows developmentally over time needs
6 to be thought about, in how you build the
7 capacity.

8 Because ultimately, states will
9 need to run it. But, you know, you have the
10 WIDA consortium as one example of something
11 that is not within any one state, but that is
12 a body that works with a large number of
13 states: 22 states.

14 MS. WEISS: Lorrie?

15 MS. SHEPARD: I guess the most
16 important point is to realize that subject
17 matter specialists should be at the center.
18 This is something where we are profoundly
19 different from other countries. Where math
20 experts come together and say what should
21 students learn?

22 What do we already know about the
23 best ways to enable that, support it. What do

1 teachers need to know? And if you put this
2 panel, constituted of math experts, you would
3 have a different flavor already of what the
4 conceptualization of the assessment should be.

5 The psychometricians need to be
6 there, because we have a higher standard for
7 equating accuracy for scaling across grades,
8 et cetera, than is possible, than has been
9 achieved in other countries. But
10 psychometricians and policy experts should not
11 be at the center of this organization.

12 So I have said a couple of things
13 about politics. We said a couple of things
14 about psychometricians. I want to also say,
15 be wary of the things that have happened so
16 far with the database decision-making movement
17 in this country motivated in large part by --
18 policy researchers do this.

19 So the blame needs to be shared,
20 and by vendors. It becomes about scores and
21 not about insight and content. And these are
22 really fundamental differences in what gets
23 produced. So who is at the table really

1 matters in these -- in putting together and
2 then also monitoring over time.

3 In other countries, the subject
4 matter experts get their results back. And
5 they say, uh oh. And they fix the assessment
6 or they fix the instruction or they work on
7 the curriculum, because too much is at one
8 grade. And that is part of what we are
9 talking about, when we are talking about
10 coherence.

11 So it should be in your thinking
12 about the governance of the consortia. And it
13 is all new to us. It is all new to us in the
14 United States.

15 MS. MELENDEZ: Lorrie, what will TC
16 practitioners claim in these groups, in
17 development groups?

18 MS. SHEPARD: Well, when I say
19 experts, I am including stodgy professors of
20 math education, and classroom teachers who are
21 teacher leaders. And in one of my slides, I
22 made the distinction between the learning we
23 should be providing for the non-expert

1 teachers, and revering the experts of the
2 teachers who know lots more than we do. So
3 there are plenty of content experts who are
4 classroom teachers.

5 MS. WEISS: Yes. Go ahead. Gary.

6 MR. PHILLIPS: I would agree with
7 everything that Lorrie just said. However, I
8 would add to it that in addition to having
9 local teachers and experts, we need
10 international teachers and experts.

11 Because you look around the world,
12 we really have our head stuck in the sand. We
13 really don't understand or appreciate what is
14 being done in other countries and how
15 beneficial it would be to us. We don't really
16 understand how far behind we are. I mean,
17 hugely behind.

18 Because we just don't have the
19 radar to see that. So it needs to be
20 international in nature. I agree, it should
21 be run by the content people; they should be
22 in the driver's seat. We need to have some
23 good international input into it.

1 MR. HAERTEL: I had, very quickly,
2 a slightly different take on your question
3 about what should you ask for in an RFP and
4 how would you recognize a promising
5 consortium. One of my points was key players.
6 And Lorrie has sort of spoken well to that:
7 the right kinds of actors at the table.

8 Another aspect of that is that it
9 is clear that the people in the states who
10 really are the decision makers are there. You
11 might consider whether or not to ask for some
12 kind of charter or bylaws or compact or
13 something in advance, so it is clear that some
14 of the governance at least had been thought
15 through and there was a sign-on to some
16 rudiments, so that you know that the states
17 that are buying in were committed to following
18 through.

19 You might also ask for some prior
20 history to show that the states have been able
21 to collaborate on some other kinds of issues
22 in the past. And you might ask for some
23 evidence that they had done their homework;

1 that they could actually discuss in a coherent
2 way what the commonalities were among their
3 philosophies or their current perspectives or
4 systems, as a place to build from.

5 MS. WEISS: I thought that was a
6 good answer for a psychometrician.

7 MR. HAERTEL: Thank you.

8 (General laughter.)

9 MR. PHILLIPS: Finally, validation.

10 MS. WEISS: Other thoughts on
11 consortia?

12 MR. LINQUANTI: Well, I keep
13 wondering about how it is going to be. When I
14 think of certain states, and the politics
15 around the adoption of standards, let alone
16 curriculum, and there are states we can all
17 name, what that is going to be like.

18 I mean I think what Ed is saying
19 about this commitment up front to seriously
20 that the governance folks, the folks in power
21 really understand what they are committing to.

22 And I worry, because what happens when those
23 folks change office?

1 Do we get an erosion or a sliding
2 back, or a distortion or a trying to change
3 gears too quickly in a different direction.
4 So I don't know how to handle that. Except
5 that I know it is going to be very
6 challenging, because there is so much politics
7 around standards adoption processes and
8 curriculum adoption processes.

9 And in a way, we want to get past
10 that, and get to what Ken and Lorrie have been
11 talking about, which is really get at the
12 underlying heart of the learning. And at the
13 curricular units. And what is really going to
14 matter. But I just don't know. I throw it
15 out there, because I don't quite know a good
16 answer to it.

17 MR. PELLEGRINO: And this is
18 somewhat less about the consortia per se, but
19 it is something that I think needs to be part
20 of what the consortia take seriously and
21 document, which is a way in which they
22 document and study and reflect on their own
23 process, because in essence what you are

1 asking people to do is to engage in a whole
2 new design process.

3 We do not fully understand, and
4 from which we need to learn. I think about
5 the experience we had with the -- going
6 through the redesign of AP science, and the
7 extent to which I wish we had further
8 documentation of all the arguments, the
9 issues, the tradeoffs, the design decisions
10 that were made as content experts and teachers
11 and learning scientists debated what really
12 should stay in or not.

13 And including things like getting
14 down to voting content off the island, which
15 they literally did at times. You come down to
16 things like that.

17 But having a way to record that,
18 and study that process, because in fact it
19 gives you, it actually gives you something
20 that is very valuable as you go forward, which
21 is, you have a way to carry forward why you
22 did what you did as people change, as regimes
23 change. As players change.

1 MS. SHEPARD: Well, just back on
2 the point of how consortia could learn from
3 each other. That is, share some of this, and
4 even create a venue for it. I will remind you
5 of the model of when NAEP started. I am
6 probably the only one that remembers this,
7 back in 1970. But the Education Commission of
8 the States began hosting conferences.

9 And one of the expectations and
10 charges of those conferences held in a Boulder
11 hotel were that the states would learn from
12 the technology that was being marshaled to
13 launch NAEP. And that was really true. There
14 were certain states that ended up
15 "piggybacking" on NAEP.

16 And there was a conference and
17 actually, a community of people that grew up
18 around that. And so that is a nice device for
19 the Department to think about.

20 MS. WEISS: Great. Let me just
21 switch topics back to some of the content
22 questions that we had. So one of the things
23 that you know, when Gary started and said, he

1 is going to disagree with what everybody else
2 has been saying, that got our attention.

3 And so but one of the things that
4 we wondered as we were listening to this is, I
5 think when we think of computer-adaptive
6 testing, we all think of sort of the current
7 multiple choice items delivered more
8 efficiently. And so I think that is one of
9 the reasons that we see things like
10 performance tests and other larger tasks as
11 being incompatible with computer-adaptive
12 testing.

13 And I am wondering if we can just
14 talk more about that, and help us see a vision
15 of where the state of the art really is, and
16 what kinds of things we might be able to do
17 computer adaptively that we are not thinking
18 about today, and what it means in terms of
19 item types and just different ways we might be
20 able to get the benefit of both of these
21 things in a system.

22 MR. LINQUANTI: Well, I think the
23 policy makers' understanding of computer-

1 adaptive testing is about ten years behind
2 what it actually is. As I said, as we speak,
3 in Oregon today, there is a huge number of
4 teachers sitting down in a large room, writing
5 items that will be administered and scored by
6 computer; instruction response items. Hawaii
7 is about to do the --

8 MS. WEISS: But tell me more about
9 that. So that is just a general statement, at
10 least fill in the blank. Are they ten-page
11 essays? What are they?

12 MR. PHILLIPS: No. They are
13 science items, conducting experiments, math
14 items with graphs and charts.

15 MR. BOND: Moving things?

16 MR. PHILLIPS: Things that will
17 move around. There are -- there is software
18 available to do writing assessments.
19 Administered and scored by computer. Now
20 admittedly, all things you like to test will
21 not necessarily be -- you will not necessarily
22 be able to administer it.

23 In most cases, you can administer

1 it. You may not be able to score it by
2 computer. And even in those cases where it is
3 scored by computer, you always have humans,
4 teachers do some validation studies. So
5 humans are always involved.

6 If you have an assessment system
7 where you have a lot of types of items that
8 cannot be scored or administered by a
9 computer, then you don't do it by computer.
10 You do it outside the computer. But you can
11 still have a computer-adaptive system that is
12 integrated with that. It just will cost more,
13 and take longer.

14 Now when I say, a score by a
15 computer, I mean those constructed response
16 items using a four-point rubric, just like you
17 would in any constructive response item is
18 part of the adaptive nature of the system. In
19 other words, those items are also used
20 adaptively. And another thing ten years
21 ago --

22 MS. WEISS: Sir, say more about
23 that. What does that mean?

1 MR. PHILLIPS: Well, what that
2 means is, the way -- there is another I think,
3 misunderstanding about computer-adaptive tests
4 among policy makers. For some reason, they
5 think that computer-adaptive testing means
6 that you are giving a break to low-achieving
7 students.

8 But that is not -- in fact, what
9 you are not doing, is you are finally giving a
10 better measurement for low-achieving students.

11 It is exactly the opposite of what many
12 policy makers mean or what they understand.
13 So to your question, what that means is, when
14 a student takes a test, let's say, I like to
15 use eighth grade math to take the test.

16 In a paper and pencil test, you
17 might get 51 or 60 items. Everybody takes the
18 same items. In a computer-adaptive test, it
19 quickly homes in on what your level of ability
20 is in far fewer items.

21 It can use items that are
22 constructed response items, just like you
23 would in a regular constructed, like a regular

1 paper and pencil test. It will do it in less
2 than 51 items. And it can do it and still
3 meet the test blueprint.

4 So for example, if you have 13
5 percent measurement items that you want to
6 have on the test, then 13 percent of the items
7 that you are administering are measurement
8 items, so it always meets the blueprint. It
9 can also do it for without going out of level.

10 Because computer-adaptive tests can
11 go way down within a grade, and way up within
12 a grade. You can cover the full range.
13 Probably most important is it can do that at
14 the student level where it meets the blueprint
15 in less time and on fewer items and more
16 accuracy.

17 It will be more accurate.
18 Particularly for the low-achieving students.
19 In fact, the measurement error for low-
20 achieving students will be as small as it is
21 for the ones in the middle. Okay.

22 In addition to that, while it is
23 doing that at the student level, across the

1 classroom, it can cover all of the
2 subobjectives. So let's say that you are
3 looking at mathematics in grade 8 and there
4 are five strands; algebra, geometry, et
5 cetera.

6 But there are 30 or 40
7 subobjectives like the Pythagorean theorem,
8 and other things like that. Well, while it is
9 assessing at the student level, every student
10 in the class adaptively, the algorithm can
11 also make sure that all of the items in the
12 content domain are covered in that classroom.

13 So that forces the teacher to cover the
14 entire content domain.

15 MS. WEISS: So it is matrix
16 sampling across all of the kids as well as
17 doing sampling?

18 MR. PHILLIPS: I wouldn't call it
19 matrix sampling. What it does, it is
20 maximizing the content coverage across the
21 classroom, while it is minimizing the
22 measurement error for the student. And that
23 is easy to do. It is already being done.

1 This is not like new technology.
2 This is existing technology. And every day,
3 different testing companies are working on
4 improved ways to provide accommodations,
5 improve universal design, improve ways of
6 measuring more and more construction response
7 items.

8 Building in performance assessments
9 into it, where you are using it, and you are
10 doing an experiment or something. So it just
11 keeps getting better and better and better.
12 And that is without seed money. You know, you
13 throw a bunch of money into this system and
14 make this the model that you want to
15 encourage, I can guarantee you that it is
16 going to ratchet up quickly.

17 And again, the final thing again,
18 in terms of scaling up, you just can't do it
19 with paper and pencil testing. You can't add
20 it up. But this, a computer doesn't care
21 whether you are testing 10,000 kids or 10
22 million kids. It is the same thing.

23 MS. WEISS: Jim.

1 MR. PELLEGRINO: If you expand the
2 notion, too, to more complicated kinds of
3 tasks. And I think Ed may have alluded to
4 this. There are more complicated measurement
5 models that we can also build which are
6 oftentimes built off of, you know, Bayesian
7 estimation routines which can also be
8 adaptive.

9 So if you take it out of the
10 context of -- we have to sort of do it for,
11 let's say an accountability or a summative
12 purpose. And we are broadening the extent to
13 which we are really trying to get a good
14 estimate, let's say in a classroom or over
15 time, where a kid is, you can essentially
16 build highly adaptive systems that are built
17 off of much more sophisticated underlying
18 conceptions of what the competence is, and
19 monitoring and tracking that.

20 But again, and you can build those
21 so that they also have diagnostic capability,
22 because the way it adapts now, is to
23 essentially do a diagnosis, rather than do an

1 estimate of skill level.

2 So there are very sophisticated
3 ways to think about this, some of which are
4 just adding more powerful psychometric models
5 onto existing tools, and putting them on
6 computers. In other cases, it is building out
7 the infrastructure for item types and the
8 measurement models.

9 MR. PHILLIPS: Speaking of
10 diagnostic tests, the computer can do a better
11 job generally than humans can do, because once
12 they know generally where your level of skill
13 is, let's say in mathematics, if you want to
14 have a diagnostic test, let's say a teacher
15 wants to have ten items or something, to get
16 more information. Those can be tailored
17 around where the student is.

18 So you can dig in deeper and deeper
19 into what the student knows, and it will do it
20 for you. And it is on the same scale as the
21 summative assessment. So you are getting the
22 same kind of information.

23 MR. SMITH: Let me just probe one

1 part on this. How deep is the assumption that
2 it is a single dimension that you are
3 measuring? In that case, are there
4 experiments where you are actually trying
5 multiple dimensions, because you are iterating
6 to one dimension, and then you are iterating
7 to another, and iterating to another.

8 MR. PHILLIPS: Most of the systems
9 right now assume one dimension. And generally
10 it is an item response three model. Three
11 parameter response. But as you just
12 mentioned, you can -- there is nothing to
13 prevent you from having multidimensional item
14 response through models.

15 Nothing to prevent you from doing
16 what NAEP does, like for example, in math, it
17 has five scales. One for geometry, one for
18 algebra. With a lot of items, that is easy
19 to do. In fact, the nice thing about this is,
20 the more items you have, which you would have
21 in a large consortium, the more things you can
22 do, and the better you can do them.

23 And so you can build in any model

1 you want. It can be multidimensional;
2 multidimensional IRT models are easy to do.
3 The Bayesian models, any type of model you can
4 do, not in a computer environment, you can
5 probably do better in a computer environment.

6 MS. WEISS: Go ahead, Ed.

7 MR. HAERTEL: We have all heard
8 this mantra of what you test is what you get.

9 And your point is well taken, Gary. We could
10 do a much better job of doing the kinds of
11 testing we are doing now, and computer
12 technology would allow us to do that more
13 efficiently and more accurately with more
14 flexibility.

15 But one thing, what sorts of -- and
16 lots of important things happen when you are
17 sitting staring at a computer screen. I do it
18 for days on end. But there are also things we
19 want kids to learn that don't involve sitting
20 in front of a computer.

21 And if that is the only kind of
22 testing we do, I worry about that happens to
23 things like oral skills collaboration,

1 extended learning tasks, sort of projects that
2 will unfold over a longer period of time. The
3 other kinds of things that we have talked
4 about, as far as performance assessment,
5 hands-on science.

6 There are things that we believe
7 are important parts of the curriculum. And if
8 we put all of our eggs in that one basket, I
9 am afraid those things are going to be left
10 out.

11 MR. PHILLIPS: Well, I would say if
12 you don't want to put all of your eggs in one
13 basket, don't put them all in one basket.
14 Just do those things outside of the computer-
15 adaptive environment. But it doesn't mean
16 don't do the computer adaption part. And why
17 not do like a huge portion of the assessment
18 efficiently, quickly? And then the other part
19 you don't do so efficiently and quickly.

20 MR. PELLEGRINO: Yes. I think that
21 speaks to this idea of you choose the tools
22 and the technologies that are appropriate to
23 the ends that you have and the particular

1 kinds of knowledge and skill that you want to
2 tap. That there isn't one right way to do it.

3 It is a mixture of ones. And they have to be
4 aligned.

5 But I think technology can allow us
6 to do a lot of things right now a lot more
7 efficiently and expand the range of things.
8 But it is not going to solve all the problems,
9 nor is it going to tap into all of the
10 competencies that we are interested in.

11 MR. BOND: Jim, that is what is
12 called the method tail wagging the content
13 dog. Yes.

14 MS. SHEPARD: Well, I think that
15 actually gets to the reaction that I am having
16 here, which is, this is being set up as well,
17 we will have the technology consortium. And
18 then we will have the other consortia.

19 And what I think that that --
20 especially when Gary makes a claim about size,
21 it pushes it in that direction. So you have
22 to realize that the big issue here around
23 curriculum is that adaptive testing will be

1 very different if you think through what you
2 want as Jim gave examples of, in the context
3 of a curricular decision. Would I -- I
4 support everything that Gary has said about
5 what it can do.

6 I have questions about whether it
7 can get there in four years for a large
8 consortium, solving the curricular problem.
9 And then what we have to worry about is the
10 same thing Jim said about what happened when
11 NCLB made everybody quickly have to solve
12 their cost constraint problems to get things
13 online.

14 So I do not think we are
15 disagreeing fundamentally in what the scaling
16 can do. I do worry about the implication that
17 you can, because this is what has happened
18 with the databased systems. These people
19 raked together a bunch of items.

20 They are not developmental in a
21 curricular sense, where you actually
22 understand how this and then this, and then
23 that. It doesn't mean you couldn't. And that

1 is when I hinted at, I think we can solve the
2 issue of how to use the strength of IRT
3 scaling.

4 And the substance of learning
5 progressions by picking particular items that
6 satisfy anchoring criteria on the scaling that
7 also have gone through the judgmental process
8 of making sense curricularly. The problem
9 with current IRT methods is that they anchor
10 by difficulty parameters only.

11 And that means they have got items
12 on there that make no sense being harder,
13 because they are harder sort of for the wrong
14 reasons, or they are harder for lack of
15 exposure. Or even some that are easy because
16 the distractors were stupid.

17 And so that is what I mean by you
18 cannot assume statistical models will
19 immediately lend themselves. And you can't
20 get just a bunch of items and put them in a
21 pot and thinking you are getting curricular
22 coherence.

23 MR. PELLEGRINO: Yes. This is, I

1 think, a really important point, because we
2 have to move away from what has dominated so
3 much of the thinking about assessment, which
4 is, we think in terms of items. Individual
5 items and we have, we sort of scale them. And
6 we put them into tests.

7 That we need to think about the
8 inferences we want to make about what students
9 know. And those inferences are best made by
10 designing something that gives you the pattern
11 of evidence. That usually is a set of items
12 carefully thought through, in terms of how
13 they are supposed to work together, or a set
14 of evidence that comes from a complex
15 performance in which you are looking at the
16 tradeoffs, in terms of different ways in which
17 the student went about doing it.

18 The inference is really -- what
19 really matters is the inference we want to
20 make. And that takes us back to thinking
21 seriously about what is the evidence that I
22 need to make a claim that a student knows
23 something in the way I want them to know it.

1 And that doesn't -- it seldom comes from a
2 single item.

3 MR. LINQUANTI: And I would just
4 add that I think we should not assume that
5 policy makers are going to hear that
6 distinction. We habituated them so much to
7 fast results that are psychometrically
8 scalable. That were not and they also think
9 that performance level tasks are just
10 problematic in terms of the large scale.

11 That there is going to be this
12 tendency to want to go to the least common
13 denominator, to go to the lowest level in a
14 sense. And I don't know how you reconcile it
15 with those. The more complex tasks that you
16 are saying we would need students to get at to
17 really know their learning.

18 Does that take place more at a
19 formative level and how does that scale up
20 summatively? I am still trying to picture
21 that. Or do we just not try and make that
22 happen at a large-scale effort.

23 MR. SMITH: Let's imagine that

1 these really are mastery of things that are
2 happening at the grade levels. If these
3 content standards, the common standards we
4 really are trying to lay out things that a
5 student should master, that strikes me as
6 leading you toward another kind of assessment,
7 potentially. Right. So what is the question?

8 MS. SHEPARD: Yes.

9 MR. HAERTEL: The answer is yes.
10 And it relates to something that was triggered
11 in my mind by Lorrie's and Jim's responses to
12 Gary, that with the kinds of IRT models we are
13 using now, we get mileage from getting lots of
14 independent tests, each of which confronts the
15 test taker with a very limited stimulus
16 configuration.

17 You give people little bits of
18 stuff, and they react. Another little bit of
19 stuff, and they react. Jim was talking about
20 configurational kinds of things where
21 collections of items -- those might be a
22 collection of items all related to the same
23 elaborate scenario. There is no fundamental

1 reason we can't -- we can do it with the
2 current IRT models; they're not the three-
3 parameter model, but --

4 MR. PHILLIPS: There are other
5 methods.

6 MR. HAERTEL: -- that's a lot of
7 work. And we don't really have any good
8 examples of doing that on a large scale that I
9 know of.

10 So, yes, the mastery progressions
11 are going to lead us to a different kind of
12 measurement, and to put the segmented models
13 first, we are not going to get there.

14 MS. SHEPARD: Actually, Wyoming did
15 this at one time, when Scott Marion was in
16 Wyoming, where they actually had mastery tasks
17 that you administered in high school about
18 heredity and your understanding of that topic.

19 And you administered when you thought you
20 were ready to take that.

21 So you could have these culminating
22 tasks that I would argue that you do not want
23 to get in the business of saying, well, we can

1 do all the rich things formatively, and then
2 different for the summative.

3 I think you could be more
4 constrained with the summative. And you might
5 even, if you had a shared curriculum, say I
6 know you have already done this. Now I am
7 going to ask you for some extensions of that
8 in this on-demand context, so when you build
9 them together, they have respectful
10 relationships.

11 MR. SMITH: I think that is a nice
12 idea. You can actually generalize from --
13 generalize toward something else. Terrific.
14 Let me just keep on this mastery for the
15 moment. Does that influence your thinking
16 about technology, using technology to
17 provide --

18 MR. PHILLIPS: No, it doesn't. But
19 it would change the way you chunk information.

20 For example, among the other things
21 that you would ask like the adaptive
22 assessment to do is go get sets of items and
23 maybe analyze them differently. Rather than

1 treating each as an independent piece of
2 information, you treat each as correlated
3 information. And in some cases, you might
4 have five items where it only represents two
5 pieces of information, something like that.

6 But again, that is all doable as
7 part of the programming and item selection
8 process. So whatever model you want, you can
9 go in currently. Most people have each item
10 as an independent piece of information. But
11 you don't need to have that.

12 MR. SMITH: So in this model, if
13 you really like would be to combine these
14 ideas. Right? Wouldn't it? That is take a
15 measurement of the whatever you have learned
16 to a mastery level of 90 percent or something,
17 but also add to it generalization questions.

18 Can you take this knowledge that
19 you now have that you have mastered, and take
20 it into another context, or take it wall to
21 wall other context.

22 MR. PELLEGRINO: Right. You are
23 building a different kind of an assessment

1 where you are actually building in notions of
2 near and far transferring. And you are
3 actually -- and you can build statistical
4 models for that as well, if you want to
5 estimate that.

6 What we are getting at here is,
7 there is nothing about the technology or the
8 measurement per se that precludes any of these
9 things. It really comes down to defining
10 these as the things that we actually want to
11 know, because they are the targets of the
12 intended curriculum and instruction.

13 MR. PHILLIPS: Right. And to
14 support what Lorrie is saying on that, that is
15 absolutely fundamental before you get into all
16 this adaptive stuff. You have got to get the
17 curriculum, the instruction, the assessment
18 worked out and determine what is important.
19 And none of that is precluded by using a
20 computer-adaptive assessment. A computer-
21 adaptive assessment, is where that all gets at
22 some point implemented or a lot of it gets
23 implemented.

1 MS. SHEPARD: And I just want to
2 push back a little bit at what I heard in your
3 question, Mike. It it is not -- okay, now I
4 have mastered it. Now I can go use it. It is
5 that what I mean by mastery is flexible enough
6 knowledge and the ability to use the
7 information so that my definition of mastery
8 means I can use it.

9 You know, the only problem with the
10 old Wyoming example is, was that they thought
11 they were done. You have to work against kids
12 thinking they have taken the test now they are
13 done with that. Building in. And you ask
14 them why they did it. It was to be done.

15 And what we want to do is actually
16 build in very visible ways that what we are
17 already done with comes up again and again.
18 And they see us calling on them to use it in
19 new ways.

20 MR. SMITH: Right. So you could
21 actually take that and those assessments could
22 happen as you go throughout the year. And at
23 the end of the year could be some sort of

1 putting it all together challenge. It would
2 be a generalization. It would be a really you
3 know, leap forward.

4 MS. SHEPARD: And it is not a
5 surprise, because they do that all the time.
6 That is the idea. We want to be modeling
7 instruction. I would never ask for an
8 extension on the final assessment that I
9 hadn't been doing routinely.

10 MR. SMITH: But you wanted the
11 feedback.

12 MS. SHEPARD: Exactly.

13 MR. PELLEGRINO: I mean, I think
14 this goes to what Lloyd said earlier. It is a
15 test worth teaching to, because the way in
16 which you are testing is essentially you are
17 preparing them to be able to do that.

18 MS. WEISS: Okay. So let me now
19 ask whether all the stuff we have just been
20 saying changes when you are talking about high
21 school versus when you are talking about
22 grades three through eight. Don't all jump in
23 at once.

1 MR. PELLEGRINO: I will sort of --
2 I am sort of perplexed by the question,
3 because I don't -- I see them as orthogonal to
4 each other. I think everything we have been
5 talking about is compatible with K-8 and high
6 school. It is just the difference is, that
7 when we are thinking about K-8, we are
8 thinking about, or three-eight, whatever.

9 A system where you are
10 systematically building to a target at the end
11 of eighth grade, which is something that you
12 agree, that is where kids need to be, before
13 we move them into this thing called high
14 school where there is variability in the paths
15 that they might take. But some, some
16 agreement on what the core academic areas are
17 that they need to master there.

18 But now we sort of focus down in
19 more on particular core areas where the same
20 alignment issues occur, the same issues about
21 assessment could be played out. And it might
22 be easier to play out some of them at that
23 level, because you are focused more on

1 particular subject matter areas that occur
2 over the space of a year. So I don't --
3 unless I am missing something in your
4 question, Joanne, I don't see the ICD as
5 essentially orthogonal to each other.

6 MS. WHALEN: Can I ask, a slight
7 twist to my question. Would you think of
8 anything that would be different then in the
9 three-three way assessment system versus a
10 high school system? So not just what we have
11 been talking about what should be similar, but
12 what should be different.

13 MS. SHEPARD: Well, whether the
14 curriculum is the same, is the biggest
15 difference. Once you have an identified
16 curriculum then everything we have said about
17 adaptive testing about the congruence of
18 curriculum, instruction and assessment all
19 pertained to that entity.

20 MR. PELLEGRINO: I think that the
21 difference is that you are also in that three
22 through eight, you are building something with
23 a longitudinal perspective of what you are

1 building.

2 MS. WEISS: And that was really the
3 nature of my question.

4 MR. PELLEGRINO: And that is a
5 foundational part.

6 MR. HAERTEL: I would say it does
7 change things. At one point, Gary said that
8 if you don't have vertical scales you are not
9 really doing growth models. And there is some
10 truth to that. I don't mean to sound a
11 negative point, but that is a valid point. It
12 is not the only position; there's twelve
13 arguments.

14 MR. PHILLIPS: Just thought it was
15 either true or not true.

16 MR. HAERTEL: Oh, it is like --
17 everything is true. In high school it is very
18 hard to say what the appropriate pretest is
19 for first course in the subject area, of the
20 assessments. Much more organized, rather than
21 course examinations. And that changes the way
22 that we model achievement.

23 It has strong implications for the

1 way the accountability system is designed.
2 Whether the impediments to high school level
3 in, under NCLB is a requirement of a common
4 test for everybody in mathematics and high
5 school. And it just gets complicated.

6 So there also problems logistically
7 in the costs of building the test in high
8 school because kids take so many more things
9 in high school, and there is so much more
10 specialization and some sets of kids take
11 different particular topics. So the cost of
12 building tests goes way up.

13 If you need a lot more tests, each
14 which is used with a smaller group of
15 students. And you have tests in all the
16 foreign languages; you have tests in economics
17 and psychology and all the other subjects we
18 have in high school. Do you include the
19 hands-on components in all the various lab
20 sciences?

21 It all just gets a lot more
22 expensive. And high school teachers can be
23 more challenging to work with. Try to get

1 high school organizations to buy into this.

2 MR. PELLEGRINO: But don't we do
3 that already, Ed? I mean, we already have
4 people taking courses in -- they are taking
5 exams in those courses. And I mean, what you
6 are talking about is if we --

7 MR. HAERTEL: But they are not
8 using part of an accountability system.

9 MS. WHALEN: So, Ed, are you
10 proposing that at a high-school level, that we
11 retain, or that we remove this element of a
12 common assessment for math. A common
13 assessment for RLA to switch to all end of
14 course exams.

15 MR. HAERTEL: Well, this is not so
16 much about the accountability system today.
17 But I would certainly favor a system that
18 provided more flexibility and provided for
19 alternative means of meeting standards, or
20 meeting expectations.

21 So that we aren't expecting all
22 kids to do the same things, because when we do
23 that, we drive it down to the common core.

1 And we lose the opportunity to assess a lot of
2 really important learning. So I would argue
3 no, everybody should not be asked to do the
4 same thing.

5 MR. SMITH: So this gets around to
6 this issue of the college ready and career
7 ready, right. And that turns out to probably
8 be an exit exam if it gets all played out, the
9 way the people are thinking about. People
10 couldn't graduate from high school unless they
11 pass that. What would that test look like, if
12 you had your druthers?

13 MS. SHEPARD: But that is something
14 to be debated and studied. Not a flip answer
15 for me today, because in other countries, they
16 do not give generic tests. And they give
17 specialized tests. And they made a decision
18 for example. So this is just hypothetical.
19 We could decide that by tenth grade, everyone
20 had to have satisfied certain common core.

21 And then there would be branching
22 that needs this higher level, that looks more
23 for some students who are not aspiring to

1 college and are looking for career
2 opportunities to very high level -- I used the
3 example engineering certification -- I am
4 sorry -- electrician certification earlier.
5 That is what people need to be thinking about.

6 And boy, we don't -- I mean, it is
7 like we have no groundwork laid and very
8 little experience for doing that. But we do
9 want to be careful not to preclude thinking
10 about that, in either this call, or the
11 revision of ESEA. That is an important thing
12 to be solved, and addressed and maybe even
13 experimented with.

14 MR. BOND: Lorrie, this notion of
15 work readiness is that even a part of the
16 vocabulary? I'm not sure I know --

17 MS. SHEPARD: Well, they are just
18 ready to -- that is a very generic test. That
19 is almost a -- I would name the company's
20 language that has led to people talking --
21 using that specific phrase of college
22 readiness. And it is typically American,
23 because it is a very generic idea. And it is

1 different from what goes on in other
2 countries.

3 MS. WEISS: So let me ask, let me
4 just sort of go back to what you said, because
5 you said that perhaps there is this notion
6 that the common standards that are sort
7 setting a minimal level of college readiness,
8 you know, may be a four-year sort of state
9 level of college readiness might even be
10 completed by sometime in high school, tenth
11 grade, whatever. And then you start varying,
12 and having differentiation after that.

13 So that might lead you to think
14 that there is more like a three through
15 college ready set of assessments, and then a
16 bunch of end of course tests in the different.

17 Like, is there a way. Is there a system here
18 that you see evolving that takes into account
19 this common core.

20 MR. PHILLIPS: That is the way I
21 would see it. Let's just use math as an
22 example. You would have three through eight
23 and tenth all on the same scale. The tenth

1 grade assessment is a challenging assessment.

2 It is on the general scale that has
3 led up to it, and all the standards up to then
4 lead up to this proficient standard, which
5 would be predictive of college ready and work
6 ready. But in addition to that there would be
7 a lot of maybe end of course exams. Those
8 would be informed I think, for example, by
9 higher education and industry.

10 So you make sure that when you take
11 a course in chemistry or physics or calculus
12 or whatever it is, that you know what, that
13 they have had substantial and representative
14 input into what it is that a high school
15 student needs to graduate with. So it would
16 be sort of like a recurrent standard setting
17 panels where you have a representative sample
18 of teachers.

19 Here you would have a
20 representative sample of university people, a
21 representative sample of industry. And build
22 that in to the end of course exams, and build
23 it into the configuration of the temporary

1 common core assessment. So but now the tenth
2 grade sort of common core assessment is not
3 going to cover calculus and chemistry and
4 biology and all these other things that you
5 like to have covered.

6 But it will be predictive of
7 college readiness. And there, the work that
8 the national assessment is going to be
9 helpful, because they have done, they spent
10 years working on this idea of how to make the
11 twelfth grade NAEP predictive of career and
12 college readiness.

13 MR. LINGUANTI: I have a question
14 that I want to put out there, because while we
15 were talking about high school, it made me
16 think, are we, do we have an assumption about
17 kids having been in the system at least from
18 grade three, all the way on through. And for
19 kids who are entering the system in, say,
20 fifth or sixth or seventh grades, what are the
21 implications for what we are sort of laying
22 out in high school.

23 And do we just have to start

1 thinking about longer time frames, and
2 building with community colleges, does it
3 become a different structure, because there
4 are lots of kids that are coming in later, and
5 particularly the kids who are English language
6 learners.

7 A lot of the four-year time frame
8 of high school and the structure of courses
9 just is not set up to meet their needs. So I
10 don't know how that figures. But I just
11 wanted to point that out as some dilemma that
12 needs to be addressed.

13 MS. WEISS: Yes.

14 MR. PHILLIPS: In the end of course
15 exam, it is very important to do it this way,
16 and make it comparable across states with the
17 expectations, because right now, what we have
18 is, in almost every case, in many places,
19 Algebra II is really Algebra I. PreCalculus
20 is at best Algebra II. Calculus is
21 PreCalculus.

22 You know, all these things have
23 been dumbed down. They are called by a

1 different name. And we need to stop that.
2 And that is why when they go into higher
3 education they are not ready.

4 MS. WEISS: Yes.

5 MS. SHEPARD: And the certificates
6 would also, if you think about sets of
7 courses. Thing about, if you were actually
8 preparing people. You wouldn't just give like
9 a math class that is related to being an
10 electrician.

11 You would literally, you would be
12 designing around a certification program that
13 in Germany and other European countries, the
14 profession takes responsibility for. And you
15 would have very rigorous course work, sets of
16 courses that applicants would have to work
17 through.

18 MR. PELLEGRINO: But that is the
19 other thing, is that in doing that, they also
20 have career paths where it is worth their
21 while to do that, because they have an
22 occupation to go to, that is respected, and it
23 is a respected path. It is not a demeaning

1 path to go down.

2 MR. BOND: That sounds like a whole
3 lot of different tests, right?

4 MS. WHALEN: Lloyd, could you speak
5 into the mic?

6 MR. BOND: No. I am saying what
7 you just described sounds like a whole lot of
8 tests.

9 MS. WEISS: Right. But it is also
10 a sort of a guideline or blueprint for how
11 testing works at that level.

12 MS. SHEPARD: You wouldn't install
13 the whole thing. You would be lucky if you,
14 right now in this competition found one
15 consortium and one profession that said we
16 want to do what Germany has been doing for
17 decades. And we are even going to use that as
18 a model. And then you would see whether you
19 wanted to generalize that, because that was
20 such an attractive set of opportunities.

21 MS. WHALEN: So at an earlier
22 session there was a panelist who said, think
23 really hard about what will need to be funded

1 by the federal government and what will happen
2 without the federal government, if we put out
3 kind of some broad statements. You guys have
4 opinions about what will not happen without
5 this assessment funding versus what will
6 naturally happen given the policy direction,
7 the national education discourse is moving?

8 MS. SHEPARD: Well, I think Jim
9 already said that. Which is, that when you
10 have a mandate about how much testing we must
11 do, and if you look at how poor the states are
12 in terms of being able to respond to NCLB and
13 even how -- they aren't actually meeting the
14 APA standards for you know, here is the
15 minimum you have to do.

16 So you are actually making it
17 possible to try out the innovative things that
18 are being spoken about, because we have
19 evidence that the states were innovating more.

20 What NCLB did was trade number for
21 innovation. And so what you are trying to do,
22 I assume probably is keep the number where it
23 is. Hard to roll that back. But you are

1 funding innovation. That is what you are
2 accomplishing.

3 MR. PHILLIPS: I would say that to
4 kick this off, you need a lot of money, maybe
5 \$350 million.

6 And then I think, to keep it going,
7 there needs to be continued funding. I don't
8 know what that level would be. It depends on
9 how the states get themselves organized.

10 But it would seem to me a proper
11 role for the federal government would be to
12 kick it off, to continue to obviously fund R&D
13 work. To improve the assessment, make it
14 faster. Make it better. That sort of thing.

15 And provide whatever funding, certainly in
16 initial stages through ESEA to keep it going,
17 depending on whatever the structure is.

18 And I think the states with the
19 funding and with the support and context of a
20 large number of states all doing the same
21 thing, and being supported and nourished and
22 encouraged by the federal government, that
23 they will get on board and do this. And so

1 there needs to be some continued -- in direct
2 answer to your question, maybe some continued
3 federal funding for the assessments. Probably
4 not 100 percent but some amount that keeps
5 them going.

6 The states do have financial
7 issues. And it is really important I think,
8 for the states to feel like there is, A,
9 ongoing support, and, B, there is an
10 environment where other states are also on
11 board trying to do the same thing. And that
12 is the only way I can think that they are
13 going to set this high standards and raise
14 their expectations to a level where the
15 superintendent is going to get penalized quite
16 a bit when schools and districts and students
17 and parents get upset that you have a very low
18 rate of proficiency in the state.

19 MS. WEISS: Sorry, go ahead, Jim.

20 MR. PELLEGRINO: There is another
21 piece that I think we talked about earlier, or
22 maybe it was over lunch that you have to stay
23 away from and the states have to stay away

1 from, because the funds can't be used there,
2 which is, it can't go build curriculum. But
3 there, I mean, Lorrie used the example of what
4 NSF funded in the past.

5 I think one role of the federal
6 government can do to agencies like NSF and IES
7 and others is they can actually support groups
8 to put together and work on the kinds of
9 integrated curriculum instruction assessment
10 tools that are consistent with what you are
11 supporting, and the professional development
12 resources. There is going to be a lot of
13 things that are needed to help the states do
14 what this is trying to do, in terms of this
15 innovation.

16 And the federal government has a
17 role to play in supporting that, because that
18 capacity is not going to come directly. It is
19 certainly not going to come from the
20 publishers. Okay, because they are going to
21 continue to try and push what they have been
22 pushing. So I think there has got to be some
23 innovation there.

1 That is a federal government role
2 through other agencies that it has, to
3 stimulate the field, to build the kinds of
4 resources that schools and teachers are going
5 to need that are the integration. And that
6 are aligned with what you want to do, because
7 ultimately, the assessment components will
8 only work well if there are the other parts in
9 place to support it, as the demand is there,
10 for those kinds of resources. The
11 instructional resources and tools.

12 A lot of which can be supported and
13 delivered actually, and made available through
14 technology. I mean, Mike, you sort of
15 advocated for open resources and things like
16 that. So there is different roles that it can
17 play, separate from just this part of the
18 funding picture.

19 MR. SMITH: One quickie. The idea
20 about college ready imparting from the problem
21 that somebody mentioned about students
22 entering a community college is in finding
23 that they can't pass an exam, and therefore,

1 they can't take a course for credit.

2 And so in California particularly
3 and in other places as well, thousands, almost
4 millions of them over time. They would get
5 stuck and they would just drop out. They
6 would say the heck with it. It is not worth
7 it. It is true in the state colleges of
8 California. You get 45 percent or something.

9 So there is a possibility of -- I
10 mean one way to think about the college
11 readiness part is to have a more general exam.

12 An exam that is shaped something like the
13 exams that they give at the college level.
14 Maybe it would be a little more challenging
15 than the average one.

16 And to then also, not to try to sum
17 it up across a whole series of end of course
18 exams, if that costs a lot of money and is
19 initially disruptive. Any thoughts on that?

20 What is your -- I mean people are
21 arguing against more generic assessment as
22 your college ready. But unless we got the
23 higher ed community aligned with us, you are

1 not gaining very much, because they will
2 probably still have to

3 MS. SHEPARD: Well, I think that we
4 have discussed the idea of a common core
5 through tenth grade. In mathematics that is
6 more general and higher level than current
7 minimum standards, and writing. If you think
8 about what colleges and universities most
9 often assess at entry, it is writing.

10 And so we can get further down --
11 we can get away from the really generic tests.

12 You know, when the kinds of college entrance
13 tests we write now, assume no content in one
14 case, and very little content in another case.

15 So you have to answer a social studies
16 question where there is a hypothetical country
17 with a hypothetical railroad.

18 And that is how we keep the tests
19 from being curriculum generic. We can have
20 general tests of mathematical knowledge and
21 writing at tenth grade that could be useful
22 for college entrance, and then still do the
23 specialization.

1 We probably, following what Lloyd
2 said earlier, would not say college ready, not
3 college ready. We would probably want to
4 signal to students that there are different
5 levels for different pathways.

6 MS. WEISS: What we have heard in
7 regards to that is, that this isn't about
8 college entrance. That that is something that
9 colleges will set for themselves. But perhaps
10 it is about placement in remedial courses.

11 So this task, whatever it looks
12 like, might be something that the state
13 universities would accept in lieu of their own
14 placement test that says you have tested out
15 of -- you have tested into regular college
16 grade work. You don't have to take remedial
17 courses when you come. Do you have thoughts
18 about that?

19 Because there is other kinds of
20 things, we could clearly incent or not in a
21 competition like this. In other words, incent
22 that the state higher ed institutions be at
23 the table saying, yes, I will sign on to this,

1 or, no, I won't. Are there pros and cons in
2 that, what you think of it?

3 MS. SHEPARD: Well, we have the old
4 CLEP tests that ETS has created. And what our
5 experience is, is that every higher ed
6 institution usually in different tiers make
7 different decisions about what they will
8 accept, even though it is nominally supposed
9 to be interchangeable.

10 So if what you are trying to incent
11 here is, a better model of doing that, then
12 once again, you are experimenting with that to
13 see if doing that better enables students to
14 learn toward a target better. What does it
15 do? And does it make students more
16 successful. So you need a logic model.

17 What are we trying to achieve here?
18 We're trying to be clear about what is
19 needed, raise the standards. So I think that
20 is within your range of experimentation, but I
21 don't think we, again, know enough. We
22 certainly don't want to just replicate what we
23 already have. So we are going to have to find

1 some new remedies through this program.

2 MR. HAERTEL: If you want to bring
3 in the college-readiness piece, you are going
4 to need to find a bigger table, because you
5 need a lot more people at it. That may be
6 beyond what you can accomplish.

7 MR. SMITH: One other thought. The
8 end of course stuff, the exams could also be
9 used to set up a system of credit for
10 performance rather than credit for seat time.

11 Would that, from your perspective require any
12 other sort of psychometric characteristics of
13 the test? Other than what you would normally
14 require? A higher level of validity?

15 MS. SHEPARD: Use the National
16 Board as an example. Yes. I mean, if you
17 really wanted a test to stand for everything,
18 we probably just wouldn't want something that
19 you could sit for in two hours. You would
20 probably want some other demonstrations. But
21 yes. You could do that. But just don't think
22 about a two-hour paper and pencil or a
23 computer administered on demand.

1 MR. SMITH: So an end of the year
2 algebra tests that you aced wouldn't qualify
3 you to go on to second year algebra. That is,
4 you can take the first year in algebra. You
5 now aced it because you just happened to learn
6 algebra when you were in seventh grade or
7 fourth grade.

8 MS. SHEPARD: If you understand the
9 definition of algebra not to just be the
10 procedures --

11 MR. SMITH: It is a test, you know.
12 So you have to come up with a test that
13 wouldn't --

14 MS. SHEPARD: You are right. An
15 Algebra I test could be rich enough to
16 enable -- to ready you for Algebra II and not
17 make you artificially ready for Algebra II.

18 MS. WEISS: Okay. Do you have
19 something?

20 MS. SHEPARD: In the last couple of
21 minutes, can we just turn to a totally
22 different question, which we have touched on a
23 little bit this morning in some of your talks,

1 but about the role that the LEA should play.

2 So with the funds that have to be
3 passed through to LEAs, what would be
4 appropriate and productive tasks that the LEAs
5 should do with the funds that are passed
6 through to them?

7 MR. PELLEGRINO: The LEAs, one of
8 the most appropriate things to do is with
9 respect to the whole issue of shifting,
10 supporting teachers and other administrators
11 to move towards understanding the ways in
12 which assessment is supposed to function to
13 support teaching and learning. I mentioned it
14 earlier, that moving from what I call a sort
15 of the model of audit and grading to more the
16 role of assessment is to probe deeply student
17 learning for the purposes of making a shift
18 towards improving that learning. That is to
19 assist learning.

20 That is going to take time, effort.

21 It is going to take resources that with
22 examples, lessons that people can learn from.

23 I think that is very appropriate for LEAs,

1 because it is right close to where the action
2 is. And the actors who have to be able to
3 sort of put in place the kind of system that
4 we are talking about ultimately. If it is
5 going to function well, because you can build
6 a wonderful set of assessments, but if
7 individuals don't know how to modify their
8 practice to assist learning in that direction,
9 then you are not going to have the capacity
10 you need. So it seems to me, that is a good
11 way for the funds to be spent, which is at
12 that level.

13 MR. LINQUANTI: And some LEAs have
14 been trying to do this already. I mean, they
15 have been doing it for years, more or less
16 perfectly or haphazardly. But the level of
17 professional development that is needed and
18 the support and the technical assistance
19 infrastructure is significant.

20 So I think just asking LEAs to work
21 on developing curriculum units or formative
22 assessments will require a lot of guidance and
23 a lot of support I think. I don't know how

1 rich the capacity is just at a straight LEA
2 level without some sort of larger support
3 mechanisms.

4 MS. WEISS: So this is something
5 you would recommend; that the consortium have
6 some answer for how the support structure
7 works, because it doesn't necessarily have to
8 be something that the LEA is providing. It
9 could be provided by the consortium or
10 statewide, and LEA people can now get into it
11 and use their money to offset release time or
12 whatever for those folks who opt in.

13 MR. PHILLIPS: You need to make
14 sure the LEAs are an integral part of this
15 whole activity; that they are on board with
16 it, understand it, and are a part of it, and
17 have the buy-in. And that would be up to the
18 consortium to demonstrate how they are going
19 to do that.

20 MS. MELENDEZ: And also the
21 professional development that it takes to
22 change instruction in the classroom.

23 MR. PHILLIPS: Yes. Right.

1 MS. MELENDEZ: And how the plan
2 that's laid out, I presume -- because that
3 would be something that would be very much a
4 part of the plan.

5 MR. LINQUANTI: Yes, because a
6 complex task that Ed was talking about
7 earlier, and Lorrie and Jim, has huge
8 implications for what the instruction would
9 need to look like. And again, just the level
10 of language involvement and what would be
11 needed to support kids using language on more
12 sophisticated tasks is a huge -- it's not just
13 about learning assessment; it is around
14 instruction. And that is what we should be
15 doing. Right? I mean, we should be getting
16 this wash-back effect.

17 MR. BOND: Joanne, if you phrase
18 this right, it may assuage some of the -- and
19 I have seen it over and over again -- the
20 outright antipathy of typical teacher for the
21 assessment community.

22 They are suspicious of us; they are
23 sick of us, because they don't see what they

1 are doing as helpful; they only see it as
2 punitive.

3 MS. MELENDEZ: Well, I think it is
4 the link that your realignment that you spoke
5 about earlier. And how it plays down all the
6 way to the classroom level, and teachers and
7 principals have the understanding of the role
8 that curriculum and instruction and assessment
9 play together. I mean, we make the assumption
10 that everybody understands that. And I think
11 it is a critical part of our learning at an
12 LEA level.

13 MS. WEISS: Well, thanks. We are
14 out of time. Let me just see. Any final
15 words of wisdom for us, that you didn't get a
16 chance to say, and have been burning on your
17 minds?

18 (No response.)

19 MR. BOND: Good luck.

20 MS. MELENDEZ: Good luck. Yes.

21 MS. WEISS: You hear that a lot at
22 the end of the session. Thank you.

23 (Applause.)

1 MS. WEISS: Thank you so much for
2 your time and for your thoughts and just your
3 generosity in sharing your wisdom and your
4 learnings with us. We really appreciate it.
5 I hope that those of you in the audience
6 benefitted from today as much as we did.

7 We are going to take a 15-minute
8 break and come back at 3:45. Anybody who
9 signed up to be a public speaker, please don't
10 take a break; instead, come up to the front of
11 the room here, and we will give you a number
12 and tell you how this process works. And the
13 rest of you, come back in 15 minutes. Thank
14 you all.

15 (Whereupon, a short recess was
16 taken.)

17 MS. WEISS: Thanks everybody. We
18 are going to get started with the public
19 speakers. So if everyone will take their
20 seats.

21 Let me just give a couple of quick
22 messages to the folks who are our public
23 speakers. I think you all know this. But you

1 have a timer on your podium. And it will be
2 green when you start talking. It will count
3 down. And it will turn to yellow when you
4 have two minutes to go and when it flashes
5 red, your time is up.

6 So with that, let's start. And
7 please, when you start, do identify yourself
8 for us. Introduce yourself to us.

9 MR. DEHOFF: Thank you, and good
10 afternoon. I am Randy DeHoff. I am the Vice-
11 Chairman of the Colorado State Board of
12 Education. Colorado is in the midst of
13 updating our standards and assessments to meet
14 the demands of post secondary and workforce
15 readiness and the 21st century. An effort
16 that was initiated by legislation in 2008.

17 Over the past twelve months, drafts
18 of the revised content area standards have
19 been developed and are scheduled for vote by
20 the State Board next week. The design of a
21 new assessment system aligned with the revised
22 standards is also underway, and it is
23 scheduled to be completed by next December.

1 From the very beginning, this was envisioned
2 as more than just a replacement for the
3 current state assessments or the CSAP.

4 Some of us went so far as to hope
5 that the new system would include components
6 that would allow it to lead to the development
7 and use of effective formative classroom
8 assessments that directly led to improved
9 instruction. Our National Association of
10 State Boards of Education study group spent
11 much of last year studying the issue of 21st
12 century assessments. And they expressed that
13 same hope.

14 The report, Reform at the
15 Crossroads; A Call for A Balanced System of
16 Assessment and Accountability issued just last
17 month. It calls for states to move to a
18 comprehensive assessment system that extends
19 down to the classroom level. I will refer to
20 that report, as I address some of the
21 questions you have posed today, and I
22 encourage you to refer to it frequently as you
23 continue to define and refine the Race to the

1 Top assessment program.

2 I would like to address the first
3 question by quoting from that NASB report.
4 Assessments should measure applied knowledge
5 and skills with the goal of all students
6 passing, rather than constructing measures
7 that describe differences in students
8 abilities. An assessment system designed with
9 the goal of all students passing is in fact,
10 the type of assessment system most compatible
11 with increasingly sophisticated growth models.

12 Such an assessment system must be
13 designed to not just measure student learning
14 but to improve student learning. If our goal
15 as education policy makers and the goal of
16 this assessment program is to bring the level
17 of assessment up to what is required in the
18 21st century, to develop assessment systems
19 that complement 21st century standards, then
20 those assessment systems must include more
21 than summative assessments and the resulting
22 data and reports.

23 The NASB report outlines what the

1 development of such a system will require.
2 Including and incorporating multiple
3 assessments into a system of curriculum,
4 instruction and educator development that
5 focuses on effective instructional practice.
6 Defining a clear set of learning goals. And
7 the system designed must ensure that the
8 resulting information from the assessment
9 system has maximum utility for guiding
10 instruction in relation to the learning goals.

11 The LEAs will play a key role in
12 the development of this system. The NASB
13 report points out that local in school
14 performance assessments serve as a dominant
15 mode of testing in most of the high-achieving
16 countries around the world. At the high
17 school level, these countries often use a
18 combination of centralized national exams with
19 primarily open ended and essay questions and
20 locally developed tests.

21 Countries in jurisdictions such as
22 Finland and Hong Kong create banks of tests
23 that teachers can draw from that include rich

1 assessment tests for classroom use. The role
2 of the LEAs does include a voice in the
3 development of the state level assessments,
4 piloting those assessments, and the
5 development and piloting of lower level,
6 formative, and benchmark assessments at the
7 district, school and classroom level.

8 If the goal is to develop an
9 assessment system that provides an accurate
10 picture of the learning that is taking place,
11 the LEA role in that development is critical
12 to ensure the system is aligned from the
13 bottom to the top. If we accept this premise
14 of the role of the LEA, then the role of the
15 teacher goes well beyond involvement in the
16 scoring of constructed responses.

17 Teachers are central to the process
18 of developing, administering and scoring
19 school-based classroom assessments as well as
20 the development and employment of in class
21 performance measures. I am sorry. The
22 development and deployment of in class
23 performance measures can serve as robust

1 teacher development to foster teacher buy-in
2 and readiness to adopt new instructional
3 practices. Teachers should be trained to
4 administer and evaluate student work using
5 collaboratively determined criteria specified
6 for standardized rubrics and scoring guides
7 all of which should be vertically aligned with
8 the higher level assessments, content
9 standards and ultimate learning goals.

10 Finally, I would offer a strong
11 admonition. Do not as federal agencies are
12 wont to do, be overly prescriptive in the
13 requirements for this grant. No one has yet
14 developed and implemented a 21st century
15 assessment system. I believe Colorado is well
16 on their way to doing that.

17 And I believe that this grant
18 program could provide a significant leverage
19 to that effort. But while the general
20 guidelines of such a system may be clearly
21 stated in the NASB report and in the grant
22 guidelines, the details of that system are
23 still undefined. I encourage you to leave

1 enough flexibility in the grant requirements
2 to encourage proposals for different
3 approaches to solving this problem. Thank
4 you.

5 MS. WEISS: Thank you.

6 MS. CRAWFORD: Good afternoon. My
7 name is Dr. Lindy Crawford. And I work for
8 the University of Colorado at Colorado
9 Springs. But I am offering comments today on
10 behalf of the National Center for Learning
11 Disabilities; NCLD.

12 NCLD is a not for profit
13 organization founded in 1977 working to ensure
14 that the nation's 15 million children,
15 adolescents and adults with learning
16 disabilities LD, have every opportunity to
17 succeed in school, work and life. Currently,
18 2-1/2 million school age students receive
19 special education due to learning
20 disabilities. Many of these students are also
21 English language learners.

22 Ensuring that these students can
23 participate in large scale assessments that

1 produce valid and reliable results is a top
2 priority for NCLD. Our organization supports
3 the accountability components of the current
4 ESEA, particularly the expanded assessment and
5 accountability provisions it contains.

6 To that end, we have produced
7 several reports designed to inform parents,
8 educators, policy makers and other
9 stakeholders of the positive impact of these
10 accountability provisions for students with
11 disabilities. Two of these reports are titled
12 Rewards and Roadblocks, and Challenging
13 Change.

14 Additionally, NCLB produced a
15 detailed report examining the current
16 situation regarding testing accommodations for
17 students with disabilities. This report
18 revealed substantial variance across states in
19 the area of allowable test accommodations.
20 Compromising the validity of what can be
21 inferred from state test data. As author of
22 that report, I am keenly interested in the
23 issue of testing accommodations in the context

1 of a new assessment system.

2 Thank you for the opportunity to
3 provide comments on the Department's proposed
4 assessment initiative. The development of
5 common high quality assessments aligned with a
6 common set of K-12 standards provides an
7 unprecedented opportunity for equity among
8 diverse learners, including students with
9 disabilities and English language learners,
10 the topic of your hearing tomorrow.

11 The next generation of summative
12 assessments must not nibble around the edges
13 of innovation. They must, given our knowledge
14 and expertise, and the flexibility provided by
15 technology, facilitate the full and equal
16 participation of all learners.

17 To that end, on behalf of NCLD, I
18 offer the following six recommendations to
19 guide the Race to the Top assessment program.

20 Number one, require assessments to be
21 designed within innovative test delivery
22 models, particularly online delivery systems.

23 Some advantages include immediate

1 score reporting so test results can guide
2 instruction, decrease administrative burdens
3 on school personnel, increase security of
4 testing materials and more flexibility in test
5 scheduling. Additionally, online assessment
6 environments allow maximum flexibility for any
7 individual accommodations required by students
8 with disabilities or English language
9 learners.

10 Number two, require a universal
11 design, UD approach to test development. Test
12 development procedures must employ UD
13 principles from the beginning to provide a
14 more accurate measure of student achievement
15 and eliminate many of the barriers that exist
16 in traditional tests.

17 A UD approach will eliminate the
18 need for many test accommodations required in
19 traditional testing situations, allowing for
20 diverse learners to show what they know.
21 Number three, require assessments that embed
22 individual student accommodations and allow
23 student control over test environments.

1 Researchers have developed systems
2 of online test environments that provide
3 accommodations that adjust to individual
4 student preferences on demand, as well as
5 online accommodation decision-making tools
6 that increase validity. Research shows that
7 the accommodations delivered within a
8 computer-based testing environment increased
9 the consistency and the integrity of those
10 accommodations and result in improved
11 utilization by the student.

12 Number four, require states to
13 accept only research-based test accommodations
14 considered as non-standard. By non-standard,
15 we mean accommodations that influence the
16 target score or the measured construct as
17 opposed to those accommodations that influence
18 an access skill or non-measured construct.
19 Any accommodation that influences a target
20 skill or the skill measured by the test must
21 be supported by rigorous research evidence.

22 Number five, require that any
23 adaptive testing be aligned with grade level

1 standards. While online testing environments
2 hold great promise, they also offer an
3 opportunity to lower student expectations
4 through adaptive approaches that adjust item
5 difficulty based on student response.

6 Such approaches are not appropriate
7 for summative assessments used for system
8 accountability. Therefore, any computer-
9 adaptive testing developed under this
10 assessment program initiative for use as a
11 summative assessment must be aligned to grade
12 level academic and performance standards. No
13 exceptions.

14 Number six, require empirical
15 analysis of test items, including the study of
16 interactions between specific items and
17 specific student populations. Items should be
18 analyzed to ensure they do not disadvantage
19 certain populations of students in their
20 format and or linguistic complexity.

21 MS. WEISS: Thank you.

22 MS. CRAWFORD: Thank you again for
23 the opportunity to comment.

1 MS. WEISS: Thanks so much.

2 MR. BURKE: Good afternoon. I am
3 Alan Burke. I am the Department
4 Superintendent for K-12 education at the
5 Washington State Department of Education. And
6 I want to thank the Department for giving me
7 this opportunity to speak to you today, about
8 these things. You do have my comments in
9 writing.

10 I am just going to make three quick
11 points, and then be able to move on in the
12 interest of time. On the one thing, the tenth
13 grade tests that we typically have as
14 summative tests do not provide information
15 about college career readiness. They
16 basically just the basic tests, they are back
17 and forth.

18 And one of the concerns we have, as
19 we head toward the end of course assessments,
20 a lot of our students are taking those courses
21 earlier than twelfth grade. They are taking
22 the tenth, eleventh grade biology or chemistry
23 or whatever.

1 So there is a bit of understating
2 of the actual ability of these kids to get
3 out. So I think the Department needs to take
4 a look at that, and a look at the whole idea
5 of readiness, including SAT, ACT, AP tests, IP
6 tests, a comprehensive look at where our kids
7 are, as they are going, so that we can make
8 sure that the information we get accurately
9 reflects where our kids are, there.

10 Number two, you speak in language
11 of varied and unpredictable item types and
12 constructive responses and performance tasks.

13 And that conflicts, of course, with have the
14 fastest possible turnaround. We know that is
15 one of these conundrums we have to deal with
16 here, with this event.

17 The summative accountability
18 functions of tests are different than the
19 formative diagnostic components. It is a
20 dance to get through both of them together.
21 Obviously, the summative tests makes valid
22 decisions about the kinds of knowledge and
23 skills that students are requiring. That is

1 one real big function of testing.

2 And of course, the formative, we
3 need to inform instruction and move on. In
4 the haste of getting stuff done quickly, we
5 are concerned about whether the summative
6 tests are going to be accurately reflect what
7 I mentioned before. On the other hand, they
8 need to get back quickly as well, too.

9 How you do that, how we are going
10 to go ahead and get through that is an
11 interesting thing. And that all is kind of
12 exacerbated by the fact that we have these
13 budget crises in the states. In Washington,
14 we are looking at \$2.5 billion deficit this
15 year. We are going to be cutting down
16 dramatically. We think we may have to talk
17 about faster and quicker ways to get things
18 back that are cheaper and keep people out of
19 the correcting business.

20 If that happens around the states
21 here, it is going to be difficult to reconcile
22 where you are going with the federal
23 government for where we have to go with

1 states. We really do have at least two to
2 three years in Washington of looking at very
3 dim budgets. We do not have ARRA to bail us
4 out next year.

5 We are going to be cutting to the
6 bone. It is going to be difficult to not look
7 at very quick ways to get scoring back. Just
8 keep that in mind.

9 And number three on teacher
10 scoring, we really value that. It is an
11 important professional development task. But
12 often it gets into over and over into labor.
13 And when it gets into labor, then you are
14 going to be having issues of who is going to
15 pay working with the unions and whatever.

16 And so the comment about teacher
17 scoring is wonderful. We want it to be well
18 supported. But we are concerned a little bit
19 about the fact that it could get us into some
20 conversations that maybe would not be very
21 productive. Thank you.

22 MS. WEISS: Thank you.

23 MR. MACQUARRIE: My name is Duncan

1 McQuarrie. I am also from Olympia,
2 Washington. And I currently serve as
3 consultant to the council of state school
4 officers where I coordinate state
5 collaboratives on technical issues and large
6 scale assessment and with 25 states
7 represented and accountability systems and
8 reporting with another 23 states.

9 I have also co-coordinated the
10 formative assessment for students and teachers
11 project. I appreciate the opportunity to
12 address you today.

13 Your proposal is consistent with
14 the council's support of and collaborative
15 projects work on comprehensive and balanced
16 assessment systems that address summative,
17 interim and formative purposes. One
18 assessment cannot achieve all of these
19 purposes. But a system of related assessments
20 can.

21 Both of the projects that I have
22 worked with emphasize that formative
23 assessment, also referred to as classroom

1 assessment, is best conceptualized as a
2 process, and therefore more akin to high
3 quality teaching and learning than it is to
4 any particular assessment instrument. As
5 such, we would urge you in your guidance for
6 the proposed competition to address the
7 critical role teacher pre-service preparation
8 programs and local district in-service
9 programs play in the long term development of
10 high-quality formative assessment processes.

11 I will leave copies of two CCS
12 papers that articulate how the formative role
13 of assessment is differentiated from summative
14 or interim benchmark roles. Validity is the
15 most important technical characteristic of any
16 test are correctly stated, the most important
17 characteristic of the interpretations that we
18 make based on test scores.

19 To retire the accountability system
20 we currently work within, and most likely will
21 continue to work in for the next time to come
22 emphasizes the ability to make valid
23 interpretations about the amount of change

1 that has taken place over time for various
2 subgroups of students. In your proposal for
3 the future, you extend this principle to
4 include the ability to measure change in
5 individual student's achievement.

6 This is a critical component of a
7 system that is to track students progress
8 towards the ultimate goal of being college and
9 career ready. A vertical scale and the
10 associated scores reflecting progress on that
11 scale is only valid if the content standards
12 associated and associated scores reflecting
13 the progress on the scale.

14 They are only valid if the content
15 standards underlying the assessment accurately
16 reflect the progress in learning necessary to
17 reach the end goal. Before assessments are
18 developed, to measure the progress of learning
19 reflected in any set of a common core
20 standards, it will be important for your
21 guidance to make provisions to fully evaluate
22 the quality and reasonableness of such
23 progressions.

1 The interpretations of the scores
2 from the assessments cannot be valid unless
3 the progressions in learning reflected in the
4 content standards are valid, and aligned to
5 the assessments themselves. Subsequent
6 predictive studies will also be necessary. As
7 long as it is in continued use and need for
8 AYP status, for schools prior to the beginning
9 of the year, there will be counter pressure to
10 include two or open-ended items that need to
11 be scored by human readers.

12 Eventually, we may be able to
13 employ artificial intelligence systems and
14 online testing to address this problem. In
15 the short term, the pressure will only
16 increase if you were to require open-ended
17 items and summative assessments be scored at
18 least in part by classroom teachers.

19 Involving teachers in scoring the
20 constructive responses can be a powerful
21 learning experience. However, the need for
22 additional training of consistently new groups
23 of scorers can lead to even longer scoring

1 periods, thus putting more pressure on an
2 already stressed summative system.

3 A better approach would be to have
4 an assessment system that includes interim or
5 benchmark assessments that also include open
6 ended items. Since such assessments are
7 typically given several times a year, they
8 would provide greater opportunity for teachers
9 to be involved in scoring and the opportunity
10 for many more teachers to participate in this
11 very meaningful experience.

12 My final point relates to the idea
13 of mandating the different formats of open
14 ended items to be used from year to year to
15 decrease the tendency to teach to the test.
16 This probably would help, but the trade-off
17 would be to decrease the equivalence of test
18 forms.

19 Equating test forms from year to
20 year, or within years begins with developing
21 comparable test blueprints. Introducing new
22 formats and different forms both decrease the
23 equivalence of the forms and therefore the

1 validity of the inferences we want to make
2 from those. Thank you for your time.

3 MS. WEISS: Thank you.

4 MR. YSSELDYKE: I am Jim Ysseldyke,
5 and I am Burkmaier Professor of Educational
6 Psychology at the University of Minnesota.
7 For 35 years, I have directed major federally
8 funded research centers focused on ways to use
9 assessment information to enhance the
10 competence of individual students. My
11 research is targeted specifically at improving
12 results for struggling students, especially
13 students with disabilities, and for struggling
14 schools. I wish to make four points.

15 First, I think there is a
16 fundamental disconnect between the stated
17 purpose of the proposed assessment development
18 activity and the criteria specified for the
19 assessments that are to be developed. Second,
20 we need a balanced assessment system that
21 includes ongoing progress and monitoring and
22 informative assessment in addition to proposed
23 summative assessments. Third, we should take

1 advantage of advances in technology enhanced
2 assessments.

3 And fourth, formative and summative
4 assessments should consist primarily of
5 multiple choice items. In my remarks, I am
6 going to focus on things that can be dealt
7 with now. I would like us not to get into a
8 long academic exercise. And I think we need
9 to focus on those over the next two years,
10 using activities that we now are aware of.

11 First, the disconnect. The stated
12 goal of the proposed assessment activities,
13 that the information gathered should be useful
14 in influencing teaching, learning and program
15 improvement.

16 The desired end of the assessment
17 development process is improved educational
18 outcomes for all students, including students
19 with disabilities and English learners. There
20 is a well confirmed knowledge base, of a many
21 important development -- of many important
22 components of effective instruction that must
23 be in place to attain improved outcomes.

1 And there is consensus that the
2 three most important are is instructional
3 match, academic-engaged time with extensive
4 relevant practice, and third, ongoing progress
5 monitoring and use of the data to adapt
6 instruction. Assessment plays a crucial role
7 in each of the above. Yet the Federal
8 Register announcement calls for development,
9 implementation of summative assessments.

10 There is a widespread consensus in
11 the professional literature and in the
12 assessment community that summative
13 assessments do not and cannot inform
14 instruction. They are not intended to do so.

15 They serve an accountability purpose.
16 Summative information is too little, too late
17 for making important differences in
18 instruction.

19 Second, the need for a balanced
20 assessment system that includes progress
21 monitoring in the formative assessments. What
22 teachers and administrators most need is
23 information-varying instruction.

1 The information that will enable
2 them to make adjustments, the kinds of
3 assessments that contribute most to
4 instructional improvements are those that we
5 used to call mastery measures and now refer to
6 under the broad umbrella of functional
7 assessment. The approaches all entail data-
8 driven decision making and are now an
9 important part of the very important framework
10 for school improvement called Response to
11 Intervention.

12 Teachers have an enormous task in
13 matching instruction to student skill level,
14 and to providing differentiated instruction to
15 today's diverse group of students, yet this
16 can be done, using computer-adaptive tests,
17 pinpoint skill level, and a match level of
18 instruction to individual skill level and use
19 an existing program, progress monitoring and
20 instructional adaptation procedures.

21 Third, take advantage of advances
22 in technology. Advances in technology now
23 enable us to do more efficient effective job

1 in assessment. To improve instruction, we
2 need more interim benchmarking and formative
3 assessments.

4 At the same time, we need to reduce
5 the assessment burden on teachers. The only
6 way that these two goals can be met is
7 computer testing and particularly, computer-
8 adaptive testing. Existing low cost
9 assessment technology already in use in
10 thousands of schools is helping teachers
11 gather more information in less time, and make
12 instructional decisions, that our research at
13 the University of Minnesota consistently shows
14 improves instructional outcomes for all
15 students.

16 And my fourth point, multiple
17 choice items have been shown to be reliable,
18 valid, inexpensive, short and able to test
19 critical thinking skills as well as or better
20 than other types of items. Yet, there is a
21 history of attempts to do away with them, and
22 replace them with constructive response items.

23 And the history of such efforts is not great.

1 They also are not amenable to the
2 accommodations that are required for accurate
3 assessments of students with disabilities.
4 The goals for development of innovative
5 assessments, which is fine. We need
6 innovation. But more importantly, we need to
7 implement interim progress monitoring, speedy
8 and informative assessments now.

9 Students can't wait. Teachers
10 can't wait. And most certainly, elected
11 officials can't wait five to ten years for
12 supposedly better assessment technologies to
13 be invented. If we want anything that will
14 improve student achievement in the next few
15 years, there is really only one way to do
16 that, and that is to use multiple choice item
17 types and technology-enhanced progress
18 monitoring interim informative assessments.
19 Thank you for the opportunity to testify.

20 MS. WEISS: Thank you.

21 MR. LAZAR: Good afternoon. My
22 name is Johnny Cash. Now, that means a lot
23 less than if I said, good afternoon, my name

1 is Clifford Lazar.

2 So I will tell you a little bit
3 about who I am. I did the 1970
4 reapportionment in the State of California
5 adopted by the Supreme Court. I did the
6 peacetime reconversion plan for Senator
7 McGovern when he ran for President.

8 I was a manager of information
9 services and management science at Atlantic
10 Richfield. And let's see -- what else. Okay.

11 I worked at three school board elections,
12 including one of my own. And I lost, but I
13 got more votes than Muskie did in Florida.

14 Okay, the purpose of assessment is
15 adaption, adapting better programs, and
16 accountability. And this is -- by the way, if
17 you want a copy of this -- and I revised it
18 right up to about a half an hour ago -- fair-
19 ed-assessment.com -- you can get a copy of it.

20 Or if you send to my email address,
21 I will send it to you. Or if you give me your
22 card, I will send it to you. Okay. Now, if
23 you Twitter, this is a tweet. It is 140

1 characters long approximately. And it says
2 what we ought to be getting out of today.
3 Okay.

4 First of all, educational
5 assessment must be done right. Secondly,
6 demographics based assessment is meaningful
7 and actionable. And value-added is not
8 demographics based assessment. Demographics
9 based assessment is superior to value-added.
10 Okay.

11 Here is an example of why you want
12 to do a good assessment. If you do a bad one,
13 and you apply it to teachers and principals,
14 they will sabotage the system, and they will
15 game the system, and that hurts students. If
16 you do a fair one, and a good one, teachers
17 will be rewarded for working hard. And that
18 benefits students. Okay.

19 Now this is the terrible math that
20 I had to do to find this out. Okay. This is
21 the result of the multiple regression
22 analysis, in which I took all of the, or most
23 of the demographics that teachers say, yes.

1 You are not thinking about the demographics.

2 This doesn't show up very well.

3 Okay. Percent African-American and percent
4 Latino, percent white. I didn't do Asian
5 because I ran out of typing fingers. English
6 learners economically disadvantaged and moving
7 in and out during the year.

8 And then I ran the multiple
9 regression against 20 high schools and the
10 only district. And I got a phenomenal, for
11 those of you statisticians, a phenomenal R
12 squared. Usually, when you do sociological
13 data, you get .3, .4, .5. I got .95, which
14 says that this demographic data does a
15 wonderful job in projecting performance at
16 schools.

17 Now let's go back. If you look in
18 the green column on the left, that is the
19 actual performance scores at the L.A. school
20 districts. If you look at the one next to it,
21 it is projected, using the math that came out
22 of the multiple regression. And we are very
23 close. Within a point or two in each case.

1 Okay.

2 Now if you take the differences,
3 that says that the school is performing well
4 or not. Okay. If the school doesn't do as
5 well as we project, something is wrong. If
6 they do better than we project, they are doing
7 something right, and there are heroes in there
8 amongst the teachers there. And we need to
9 find out what their practices are, and try to
10 spread them around the system.

11 And we got As, Bs, Cs, Ds, and Fs.

12 Now interestingly, just below the blue line
13 is the best performing high school in Los
14 Angeles. And we gave it a C, because they
15 actually underperformed.

16 That means that the teachers there
17 are sliding on their demographics. Whereas,
18 schools that did much poorer who got As, okay,
19 they were overperforming what they should be.

20 And those are where the heroes are. And
21 those are the ones who should be rewarded.
22 Okay. Here is a table which compares value-
23 added.

1 Now I want to jump to the last
2 slide. Okay. I didn't get time to do crime.

3 This is crime in Los Angeles. Monroe High is
4 in a nice district. Jefferson High is in a
5 poor district. These are crimes committed in
6 seven days within five miles of the schools.

7 Obviously, crime is an issue. And
8 teachers know it is, and we have to take it
9 into consideration. Thank you very much for
10 your time.

11 MS. WEISS: Thank you.

12 MR. NATALI: Good morning, ladies
13 and gentlemen. Hello, Board. What I want to
14 talk to you about today is an assessment that
15 I did on my Department of Business Studies.
16 At the beginning of the semester, I walked in
17 the office at the wrong time, and became the
18 Assistant Chair of the Department. My first
19 question was, how well is our program working?

20 How are we doing? So I looked at some of the
21 details. Thank you.

22 I looked at some of the details
23 about Pikes Peak. We are the second largest

1 college, or second largest community college
2 out of the thirteen in Colorado. And what is
3 interesting is that our student population is
4 larger than the University of Colorado at
5 Colorado Springs, Colorado College, and the
6 Air Force Academy combined.

7 My goal was to see if there is any
8 overlap in the knowledge of freshmen and
9 sophomores. Since we are a community college,
10 we have beginning freshmen, graduating
11 sophomores. Okay. And everybody in between.

12 If there is an overlap, that means people are
13 falling through the cracks, and we don't want
14 that. So my assessment was designed to make
15 sure that there was no overlap, or actually,
16 to see if there was an overlap. Let me put it
17 that way.

18 The criteria I used was, basically,
19 they had to be registered in any one of our
20 business programs, either an AA, AS or a
21 certificate. We considered a freshman to be
22 in the first semester, less than 18 credit
23 hours. And we considered sophomores ready for

1 graduation in their last semester of more than
2 44.

3 I looked at just the seven classes
4 I teach this semester. I have 129 students.
5 Forty-six were qualified in those two
6 criteria; 23 responded, and 14 actually took
7 the assessment. Pretty low turnout, but it is
8 a pilot program, and hopefully we will get
9 this going on in all the other departments.

10 This is an example of what the
11 students faced. What we did was we took all
12 the disciplines. We took economics,
13 accounting, marketing, management, and we
14 decided to just take the top five questions.
15 That is all. Just the top five questions.

16 So I went to each of the faculty
17 and said, give me your best questions that you
18 would ask somebody a year, two years out after
19 they are out of the school. What do you want
20 them to remember. So everybody gave me a
21 couple of questions.

22 We had 30 questions total. Each
23 was worth two points. So that they went

1 online. They took this test. And my students
2 are pretty used to taking these online tests.

3 Then the other 40 points came when they had
4 to write an essay. Now they had some
5 limitations. They were only limited to 60
6 rows, it is 80 characters wide. There is no
7 spell check. No grammar check.

8 So this gave us some pretty
9 interesting essays. And so what I did was, I
10 took the essay. I copied into Word. Sent it
11 to another professor. I graded it. Another
12 professor graded it. We combined the scores
13 so that there is no bias, or very little bias.

14 My hypothesis is that they
15 shouldn't overlap. I used this equation. Now
16 there is a test on Friday. So memorize it,
17 please. Okay. So it is the student's pooled
18 T test. The results are pretty amazing. The
19 red area, if the results would have landed
20 anywhere in that red area, then we have a
21 problem. You know, they are basically a
22 freshman that knows just as much as a
23 sophomore.

1 But what happened was, the answer
2 was 18 units away from each other. So they
3 are very far away from each other. Not even
4 close. So the bottom line, freshmen and
5 sophomores are far from each other. Virtually
6 no chance that they are going to fall through
7 the cracks. If you can't measure it, you
8 can't manage it. That has been my motto all
9 along.

10 And one of the things that we did
11 learn from this, is that we need more data
12 rich studies. So I am pushing the other
13 departments to cough up some more student
14 volunteers and hopefully, we can come up with
15 a deeper, richer data. And that is all I have
16 got.

17 MS. WEISS: Thank you.

18 MR. NATALI: Thank you.

19 MS. ZELMAN: I am Susan Zelman. I
20 am presenting testimony today as a former
21 superintendent of public instruction for ten
22 years in the State of Ohio, and now Senior
23 Vice President for Education for the

1 Corporation of Public Broadcasting, where I
2 serve as chief education policy advisor and
3 consultant to the public service media system.

4 It is my pleasure to provide
5 comments to the U.S. Department of Education
6 regarding the proposed Race to the Top
7 assessment initiative. The Corporation for
8 Public Broadcasting is a private, non-profit
9 corporation that was created in Congress, by
10 Congress in 1967. It promotes universal
11 access to public telecommunications services
12 by supporting over 1,100 radio and television
13 stations across America.

14 The Corporation for Public
15 Broadcasting has a long and well documented
16 record of funding for diverse and innovative
17 educational programming that is second to
18 none. However, beyond programming, public
19 service media helps teachers, caregivers,
20 parents and communities educate children. CPB
21 is a strong ally for raising the academic bar
22 and closing achievement gaps for all students,
23 particularly the under represented and

1 underserved.

2 Because I have been a policy leader
3 in both public education and now public
4 service media, I understand how publicly
5 funded television and radio stations can
6 enhance a national system of student
7 assessment. And a national system of student
8 assessment can provide for better integration
9 of curriculum instruction, assessment, and
10 educator development. And public service
11 media can provide the digital content and
12 technological know how to assist the
13 innovative and digitally based system.

14 As state superintendent, I wanted a
15 coherent, comprehensive assessment system that
16 assured that all students have the opportunity
17 to learn. In Ohio, we saw a teacher
18 assessment system that was built upon clear
19 and distinct academic content standards that
20 incorporated 21st century skills such as
21 problem solving, innovation and collaborative
22 learning.

23 We envisioned a performance

1 assessment system embedded in many curriculum
2 units that would be crafted to teach teachers
3 how to individualize learning opportunities
4 for students through individualized student
5 plans. As state superintendent of public
6 instruction, I saw how the funding constraints
7 limited Ohio's important opportunity to
8 develop formative and summative assessments.

9 Therefore, I got a grant from the
10 Gates Foundation and the Hewitt Foundation --
11 thank you, Mike -- for working with Stanford
12 University in 24 school sites to develop the
13 performance assessment task with strong
14 statistical validity and reliability systems
15 modeled after the moderation panels found in
16 Queensland, Australia, Finland and other high-
17 performing countries. To show the value of
18 public service media on a national assessment
19 system, I would like to address technology and
20 innovation in assessment and project
21 management in consortia.

22 Public service media has a rich and
23 trusted digital content such as video and

1 audio programming, online games, simulations,
2 podcasts, and other digital learning objects,
3 which will allow for multiple representations
4 of similar concepts essential for teaching and
5 assessing a diverse group of learners. These
6 resources, which are much of it in the public
7 domain, motivate and engage students. Some of
8 the content has been subject to rigorous
9 evaluations, and demonstrate the efficacy for
10 enhancing learning outcomes for poor and
11 underserved children.

12 Our public service media is in the
13 process of aligning this content with academic
14 standards through the Digital Learning Library
15 and our American Archives project. In
16 addition, we have customized digital learning
17 objects for assessment projects. These
18 resources can be used for performance
19 assessment tests, performance portfolios,
20 constructed responses, and essays.

21 The Public Broadcasting System has
22 a long history of online educator professional
23 development funded by the U.S. Department of

1 Education. Systems like PBS Teachers Online,
2 ThinkPoint, New Learning for Educators and
3 Teachers on Main are but a few examples. Our
4 system is and can be even more helpful for
5 facilitating teacher and administrative
6 training in assessment literacy.

7 Specifically, our system can
8 provide online training in developing items
9 for formative and summative assessments,
10 scoring to performance tests, the
11 interpretation and use of the results of all
12 types of assessments, the creation of
13 curriculum materials that can be shared
14 electronically within and across districts and
15 schools. Public Broadcasting System is now
16 experimenting with new digital media such as
17 iPod, cell phones, mobile TV and other
18 handheld devices that can serve to help test
19 developers as they continually readapt to new
20 and ever changing technology.

21 We are, in fact, a leader in
22 adaptive technologies, for example, we have
23 great success with WGBH in Boston. We have

1 content in Spanish. And we believe that we
2 are a valuable partner in a multiple
3 consortium which can in fact, help create a
4 new vision for public service media and
5 assessment working together.

6 MS. WEISS: Thank you.

7 MS. ZELMAN: And we are in the
8 public domain. Thank you very much.

9 MS. WEISS: Thank you.

10 MR. KAHL: Hi. I am Stuart Kahl,
11 CEO of Measured Progress. We are an
12 educational testing company that has been
13 around for 26 years. We have done nothing but
14 statewide large scale assessment programs in
15 alternative and general assessments. And I
16 want to thank you for the opportunity to speak
17 today.

18 I have three points. I talked to
19 Lorrie during the break, and told her that one
20 of these days, I am going to come up with
21 three points that she doesn't make before I
22 do. But I am going to make them anyway. They
23 are important points, I think. And things

1 that you are obviously debating right now.

2 One has to do with the size of the
3 consortia. We are the contractor for NECAP.
4 We have been the contractor from day one. We
5 worked in those states for many years before
6 that, in their assessment programs. And so I
7 can tell you something about what it takes to
8 make a group of states work together. And it
9 is a challenge.

10 I would say that to favor or
11 encourage large consortia would be to ignore
12 the factors that made NECAP a success. It is
13 clearly something that is a function of
14 relationships. There is a big difference
15 between policy decisions to accept a set of
16 standards, or a policy decision to accept a
17 common test or to accept an existing system.

18 And when you want people to roll up
19 their sleeves and create a new innovative
20 system, that is something else again. And as
21 I said, it is all about relationships. It is
22 about ownership. It is about influence of
23 each state. There is some input that counts

1 and is meaningful so that they do develop a
2 sense of ownership in the process, and the
3 product.

4 It is a question of management.
5 The NECAP states were small like-minded
6 states, geographically close. That was very
7 important. They met frequently, face to face,
8 rolled up their sleeves, got a lot of credible
9 work done, and worked with incredible success.

10 As an aside, I might add that the
11 savings that everybody talks about with these,
12 are not the same, necessarily with large
13 states. So if you have large states doing a
14 consortium, they already have the efficiencies
15 of scale that the smaller states accomplish by
16 banding together. Those are related to the
17 variable costs that are student based, based
18 on the number of students, like scoring,
19 shipping, receiving, things like that.

20 So the larger states would not gain
21 a benefit of tremendous savings there. The
22 same thing with the fixed costs and things
23 like program management and test development.

1 Those are for very large states, those fixed
2 costs almost negligible compared to the
3 variable costs per student costs. So in any
4 case, my recommendation is not to favor or
5 encourage large consortium.

6 Multiple measures, a good practice.

7 If I commend the Department for that
8 commitment to multiple measures. There has
9 been a lot of talk of some sort of interim
10 components, interim components of a variety of
11 types. I should mention that certain teachers
12 have often had that interview or test, and
13 they always complain, well, you are testing
14 some stuff that we didn't teach within the
15 past six months.

16 Well, I am sorry. But I hope for
17 accountability's sake that retention is a
18 concern. And so we wouldn't be teaching
19 things we want people to retain. I am not
20 just talking about low level knowledge. I am
21 talking about skills and deeper understandings
22 as well.

23 And so I think there is a place for

1 those end of the year summative kinds of
2 tests. I think they have to model good
3 testing. I do have a problem with all
4 multiple choice tests. The literature is
5 pretty clear on the negative impact of all
6 multiple choice testing in a high stakes
7 environment.

8 So in any case, I think there
9 should be some guidelines for what the interim
10 assessment should look like, what they should
11 try and accomplish, what they shouldn't
12 accomplish. I don't think they should be
13 testing just recently taught low level
14 knowledge. I think they should be tapping
15 those skills that aren't easily tested by
16 those other meaningful and important
17 components that are administered at the end of
18 the year, the other summative components.

19 Teaching, excuse me, teacher
20 scoring. There is no question there is a
21 tremendous benefit of teacher involvement in
22 the scoring process. We have been doing that
23 for years in a lot of our programs. It is

1 considered the best professional development
2 the teachers have had, very often.

3 And my only comment there is let's
4 not go overboard with it. It is like the
5 consortia, you know, they work. But we only
6 have one that we have seen with a
7 comprehensive program that has been
8 successful. It is the only one that has
9 existed. And okay, now let's have more of
10 them and make them bigger.

11 Well, it is the same thing with
12 teacher scoring. There are times when you
13 want the teacher scoring. There is time when
14 you may not. When you have a component such
15 as that end of the year summative on demand
16 component, that is the time when you want to
17 use those testing contractor systems. They
18 are good.

19 The literature, sometimes,
20 newspaper articles, the one this past, I guess
21 two months ago, one last summer blasting the
22 scoring process. Often digging into the
23 qualifications of the scorer. The people who

1 wrote those things are uniformed. And those
2 articles are baloney, quite frankly. I'm
3 being kind, I think. I could be a little
4 stronger about that.

5 But the point is, the quality of
6 the scoring is documented in the technical
7 manuals and other materials produces by the
8 contractors. Their systems are very good.

9 So my recommendation is, that the
10 teachers be involved in those interim
11 components. They still get those same high
12 quality experience of scoring and training
13 that is somewhat associated with it. And if
14 they are doing those interim curriculum-
15 embedded performance assessments, that would
16 be my recommendation. Thank you.

17 MS. WEISS: Thank you.

18 MS. PAPINI: Hi. My name is Jodi
19 Papini. And I am a teacher at Douglas County
20 school district, just down the road here in
21 Colorado. I have been teaching for 15 years
22 with students in the general classroom
23 setting, helping students succeed is my

1 ultimate passion.

2 I am a member of the Douglas County
3 federation and the American Federation of
4 Teachers. I have provided professional
5 development for teachers through the AFT's
6 ERND Thinking Mathematics courses, and some of
7 which I helped to develop. I am currently a
8 math instructional coach in my district.

9 Today, I speak on behalf of
10 teachers. I ask that as you write the
11 guidelines that will shape the development of
12 the next generation of assessments, you please
13 consider AFT's smart-testing criteria, which
14 starts with strong grade specific content and
15 standards, and includes a number of
16 interrelated pieces; well developed, grade-by-
17 grade curricula, assessments aligned to
18 content standards. And an efficient, valid
19 and reliable testing system that does not
20 duplicate testing across the education systems
21 level.

22 Appropriate inclusion of English
23 language learners and students with

1 disabilities in testing programs. Timely
2 provision of user friendly testing results for
3 teachers and students. Supportive
4 professional development, including coverage
5 of what the content standards are, and how
6 they relate to the state curricula and
7 assessments.

8 How to teach to the content
9 standards and how to use testing data to
10 inform instruction. Accountability for
11 results, and transparency of the system. Some
12 important pieces of the smart-testing criteria
13 have been clearly violated or neglected under
14 the current system.

15 Standards are often so broad and
16 ambitious, even in places that have grade-by-
17 grade curricula, the expectations are
18 unrealistic and overwhelming. There is so
19 much material to cover in a school year.

20 Some teachers have expressed
21 concern over not having the time to take
22 advantage of teachable moments. Teachers are
23 at times faced with difficult choices, like

1 taking an extra day or two to reteach the
2 material that they know students have not
3 mastered, knowing that at the end of the year,
4 they will be rushed or simply not able to
5 cover all of the required material. Under the
6 current assessment systems, states are
7 mandated to administer summative assessments
8 once a year.

9 However, some states in many
10 districts have developed additional interim
11 and or benchmark assessments resulting in
12 multiple layers of testing at the classroom
13 level. During focus groups conducted by AFT,
14 teachers have calculated that up to 20 to 25
15 percent of the school year can be consumed by
16 summative, interim and benchmark assessments
17 alone.

18 These additional assessments often
19 aim to emulate the summative assessments such
20 that students are tested and retested on
21 similar material. In other cases, these
22 assessments do not align to the summative
23 assessments so that teachers spend the school

1 year, administrating assessments and receiving
2 data that does not align or inform progress
3 towards higher achievement on a summative
4 assessments currently used to evaluate
5 schools.

6 This practice does not make the
7 best use of the already scarce instructional
8 time. In developing the next generation of
9 assessments, require that those who have
10 received the development and implementation of
11 these assessments develop a system that
12 incorporates aligned standards, curricula,
13 assessments and professional development.

14 Tests that do not duplicate across
15 educational system levels. User friendly test
16 results. Accountability of results.
17 Transparency. And appropriate inclusion of
18 English language learners and students with
19 disabilities. And most important, require
20 that they take into account the impact of such
21 assessments on the day-to-day classroom
22 experiences of our children. Thank you.

23 MS. WEISS: Thanks.

1 MS. GIBBS: Good afternoon. I am
2 Stella Gibbs. And I am here representing
3 Pacific Metrics. And we are an assessment and
4 technology company out of Monterey,
5 California. We know that the consortium must
6 design assessment systems to include several
7 appropriate measures, not just one measure.

8 And let me give you an example.
9 During the No Child Left Behind era, many of
10 us heard a great deal of complaints from
11 teachers about the time that summative testing
12 takes away from their instruction.

13 And it is clear that many American
14 teachers don't see assessment as an investment
15 in understanding what their students know, and
16 what they are able to do. Even in the testing
17 industry, we understand this criticism.
18 Summative tests that are administered in the
19 spring often don't provide teachers with
20 timely or appropriate feedback.

21 So we have to provide a more
22 authentic assessment environment, and move
23 beyond the limitations of paper pencil tests

1 and an over reliance on multiple choice items,
2 and include item types that engage students
3 and let the demonstrate what they can do.
4 Many of us would like to see the confluence of
5 instruction and assessment occur over time,
6 and some of us are taking steps towards that
7 now.

8 But the assessment system that we
9 can build today, I mean, right now, with the
10 existing technology and existing budget
11 limitations can contain four different
12 components. Diagnostic testing at the start
13 of the school year to establish general areas
14 of student strengths and weaknesses, delivered
15 online so that results are generated quickly,
16 and this could be a cut, or a fixed form test.

17 Summative testing, it is time we
18 recognize and say out loud that the real
19 benefactor of summative testing or
20 accountability systems and they have been
21 designed by policy makers. Teachers and
22 school administrators have told us over and
23 over that summative tests are not especially

1 useful.

2 And so this means that we need to
3 rethink it. If summative testing serves an
4 accountability purpose, we can give students
5 shorter tests, take advantage of matrix
6 designs and still come out with more useful
7 information for policy makers than can be
8 produced under the fixed form testing system.

9 Three, benchmark testings during
10 the year can include a significant proportion
11 of extended response items so that students
12 can demonstrate their skills. And this is the
13 way the assessment design can take advantage
14 of new technologies both for administration
15 and ultimate scoring. And extended responses
16 don't have to be just a written response.

17 Technology can allow for
18 interactive items. The point here is to let
19 students demonstrate the complex materials
20 that they have mastered and apply their
21 knowledge. Don't let them go on autopilot and
22 just check a box. Formative assessments
23 administered by classroom teachers can utilize

1 interactive items and performance tasks so
2 that students can be actively involved in
3 problem solving. I would like to leave you
4 with the following points.

5 Point one, link assessment and
6 instruction together wherever possible. They
7 are not discrete elements of the students
8 education. Embed the assessments into the
9 curriculum, into the classroom. Provide
10 results and feedback in the classroom setting.

11 And help students connect assessment with
12 their learning.

13 Point two, increase teacher
14 involvement in scoring student work, but be
15 realistic about the time they have available
16 for it. For example, you can design a system
17 that uses automated scoring for a first read
18 or base score, along with a human component,
19 or you can use teacher as a second reader or
20 as part of a professional development
21 activity.

22 Three, funding. We have heard that
23 Secretary Duncan will dedicate up to \$350

1 million of funds to develop new assessments.
2 Yes, there are efficiencies from sharing
3 development costs. But this is just one
4 component of an assessment.

5 If all states in a consortia want
6 to implement the same test at the same time,
7 then they have to design creative solutions to
8 administration, scoring and reporting. And
9 the issues currently associated with large
10 scale administrations are going to be
11 magnified and more complex than they are
12 currently. And if the states implement and
13 test at different times, test security
14 solutions must be designed also.

15 Finally, on that same address,
16 Secretary Duncan said, this is a growth area
17 for the testing industry which may worry some
18 that assessments used across multiple states
19 will be bad for business, even if it is the
20 right thing for kids. However, it is not my
21 job to worry about their business. I
22 respectfully suggest that some of the most
23 distinguished and knowledgeable assessment

1 experts have chosen to work in the testing
2 industry, and spend every day immersed in the
3 issues and practicalities of large scale
4 assessment and technology.

5 Some of them are on the panel.
6 Some of them are in this room. Some of them
7 have been appointed to the content and
8 standards committees. And others are working
9 passionately in their chosen careers as we
10 speak now.

11 So what I would like to say is,
12 don't discount their expertise, just because
13 they work for a vendor. In many cases, they
14 share precisely the same goals for educational
15 improvement as their colleagues who works for
16 government and for universities. Thank you.

17 MS. WEISS: Thanks.

18 MR. GIANNESCHI: Hi. Good
19 afternoon. My name is Matt Gianneschi. I am
20 Governor Ritter, the Governor of Colorado,
21 Governor Ritter's education advisor. And I
22 guess I have the distinction of being last
23 today. Maybe that is because I have the

1 shortest commute. I can walk after this.

2 Let me be the last to welcome you
3 to the great State of Colorado. Governor
4 Ritter had intended to attend today's hearing
5 and share a few thoughts with you, but he
6 actually unfortunately was on his way to
7 Washington, D.C. to meet with Joanne's boss.
8 So we will do that tomorrow. And so he sends
9 his regrets.

10 Anyway, on his behalf, and on
11 behalf of the National Governor's Association,
12 Governor Ritter is the Chairman of the
13 Education and Workforce Committee for the
14 National Governor's Association. Obviously,
15 this has been a big topic for those of us who
16 work in the education policy field. I am
17 going to talk just very briefly about a couple
18 of general points.

19 I won't go through my testimony. I
20 think you will appreciate that, because it is
21 late, and there is a lot of great things about
22 what is going on in Colorado. But I think Mr.
23 DeHoff covered most of those. A few things

1 that I want to make sure that are imbedded in
2 whatever comes out in the guidance, but
3 related to the use of assessments. And this
4 is a practical matter.

5 We in Colorado have been heavily
6 involved in the practice or at least in the
7 planning for a new system of assessments. And
8 so this couldn't have come at a better time.
9 However, we have realized that there are a
10 variety of obstacles and policy barriers and
11 are moving quickly to adopt whatever this new
12 great 21st century system of assessments is
13 that you help us design to some degree.

14 So to that end, we respectfully ask
15 the Department of Education to consider ways
16 that participating states, the states that are
17 selected for this process are enabled through
18 either a waiver process, or some other way to
19 relax existing strictures that are found in
20 federal statutes that might impede quick
21 progress for states. I think that there is an
22 immediacy.

23 There is an urgency that states

1 have. Policy makers want to move quickly.
2 And there may be some things that the federal
3 government can provide by way of waivers that
4 will allow the states to be creative and to
5 adopt a new system of assessments very
6 quickly.

7 Secondly, and similarly, focusing
8 on online assessments, I think you have heard
9 much of that today. And we are very strongly
10 supportive of that goal. However, I think
11 that much of this is going to have to require
12 investments in the states.

13 Here in Colorado, we have places
14 where we may have the best intentions of
15 putting together online assessments but we
16 don't have the bandwidth to reach the schools.

17 It is the last mile problem that our
18 technology officers like to talk about. We
19 are not -- that is obviously beyond the scope
20 of this assessments plan. However, you can
21 help us.

22 We currently spend upwards of \$18
23 million a year on our existing summative

1 assessment program. It doesn't mean that it
2 is a bad program; it is a good program.
3 However, if we are able to quickly adopt other
4 options, or to have investments from the
5 federal government to support the state, we
6 might be able to reinvest existing state funds
7 to support our schools, to enable them to get
8 better access to the kinds of technology that
9 are going to be necessary to make this work
10 well for their students.

11 And so consider ways that states
12 can help themselves. And we are not asking
13 you necessarily to foot that bill. But we do
14 think that there are ways, I mean, this
15 environment, this economic environment that
16 you can support us.

17 Finally, there has been a lot of
18 conversation about the idea of college and
19 workforce readiness. We call it college and
20 career. We call it post-secondary workforce
21 in our state.

22 But one thing that we would
23 respectfully ask is that states that

1 participate in this, that those state systems
2 adopt a common definition of what that
3 actually means. Right now, it is really
4 ambiguous. There doesn't seem to be a common
5 definition. We have defined it in Colorado.
6 I am not suggesting it is right. It is right
7 for us right now. And it may change.

8 But we have adopted, we have
9 collaborated between K-12 and higher education
10 to come up with a common definition. And what
11 that has allowed us to do, is it has enabled
12 the state to come up with something meaningful
13 for students so that we can provide them with
14 assessments, provide teachers with tools and
15 the information that has direct relevance to
16 the next phase of their lives, whatever that
17 might be.

18 So again, the three areas that you
19 can help us, relaxing the standards or the
20 strictures and considering waivers for states
21 that participate in this plan, helping states
22 help themselves with providing additional
23 online support, and then asking participating

1 states to adopt definitions of workforce
2 readiness. And I will yield my three seconds.

3 MS. WEISS: Thank you very much. I
4 think Matt was our last speaker. So let me
5 just close by once again thank our experts for
6 coming all this way, and sharing their
7 thoughts and their wisdom with us.

8 And thank all of you in the
9 audience. Some of you I know have joined us
10 for several of these sessions now. So thank
11 you very much for coming. I hope that this
12 has been useful to you. We look forward to
13 seeing some of you, I think, again tomorrow.

14 And again, many thanks for coming.

15 And thank you to Colorado for hosting us in
16 this beautiful location. Thank you.

17 (Applause.)

18 (Whereupon, at 4:48 p.m., the
19 meeting was concluded.)
20
21
22
23

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10