

U.S. DEPARTMENT OF EDUCATION

RACE TO THE TOP ASSESSMENT PROGRAM
PUBLIC & EXPERT INPUT
MEETING

10:00 a.m.
Tuesday,
November 17, 2009

Salon E
Atlanta Airport Marriott
4711 Best Road
Atlanta, Georgia

FACILITATOR: JOANNE WEISS

PANEL MEMBERS:

CAROL CAMPBELL
LOU FABRIZIO
THOMAS FISHER
BRIAN GONG
PAUL NICHOLS
ALEXA POSNY
THERESA SISKIND
ANN WHALEN
JUDY WURTZEL

A G E N D A

| <u>SPEAKER</u> | <u>PAGE:</u> |
|---------------------------|--------------|
| <u>Expert Presenters:</u> | |
| Lou Fabrizio | 21 |
| Carol Campbell | 50 |
| Theresa Siskind | 83 |
| Brian Gong | 109 |
| Thomas Fisher | 144 |
| Paul Nichols | 170 |
| Roundtable Discussion | 201 |
| <u>Public Presenters:</u> | |
| Roy Truby | 287 |
| Stephen Lazer | 292 |
| Jeff Hubbard | 299 |
| Joe N. Weems, Jr. | 304 |
| Gerald Eads II | 309 |

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

2 MS. WEISS: Good morning
3 everybody. Thank you so much for joining us.

4 I think we're about ready to get started.
5 Before we do, I just want to see whether
6 anybody needs interpretation services. We
7 think maybe somebody does. And if you do,
8 we've got a space for you at the front of the
9 room.

10 (Pause.)

11 MS. WEISS: Okay. Maybe not.
12 Thank you. So welcome. My name is Joanne
13 Weiss, and I am the director of the Race to
14 the Top program at the Department of
15 Education.

16 I want to start by giving you a
17 quick overview of what we're trying to
18 accomplish today -- what the purpose of this
19 meeting is and why. And then we will turn
20 most of the air time over to this fabulous
21 group of experts that we have talking with us
22 today.

23 So the first thing that I wanted

1 to do was make sure you knew why you were
2 here. The Race to the Top competition that
3 you have probably been hearing about a lot --
4 those of you in the states have certainly
5 been hearing about it a lot. The regulations
6 for that were announced last week. And that
7 is a \$4 billion competitive grant to
8 encourage and reward states who are
9 implementing comprehensive statewide reforms
10 across four key areas.

11 What we're here to talk about
12 today is not that program. It's the
13 companion program, which is a separate
14 competition for up to about \$350 million.
15 And that's designed to support consortia of
16 states who are trying to implement common
17 standards by having the Department help fund
18 the development of a new generation of common
19 assessments that are aligned to those
20 standards.

21 So that's what we're here about --
22 here to talk about today. The applicants for
23 this separate competition, the assessment

1 competition, are consortia of states. And
2 just like with the regular Race to the Top
3 program, 50 percent of the funding does have
4 to be passed through to districts in the
5 consortia. So you'll also probably hear us
6 talking about what uses of funds might make
7 sense for that money.

8 That's funding that goes through
9 to states -- to districts based on a formula,
10 but the states can direct the use of those
11 funds to be consistent with the program or
12 the application that they've put forth. So
13 we'll talk a little bit about that today I'm
14 sure.

15 And just to give you a sense of
16 the timing for this competition and the
17 process here, typically in the Department, as
18 I'm sure many of you know, we put out a
19 preliminary notice. We put it out for public
20 comment for 30 days and then we take all of
21 that comment into consideration and release a
22 final notice. That's the process that we
23 just went through with the regular Race to

1 the Top regulations.

2 We're actually doing a different
3 process for this grant. And we're doing a
4 different process for it primarily because as
5 we started at the Department to put pen to
6 paper in designing what might have been a
7 preliminary notice we really realized that we
8 at the Department -- and we felt like it was
9 probably also true of many of you in the
10 states -- could benefit from a lot more input
11 at the front end before we sat down to try to
12 define what this might look like.

13 And so we worked with a number of
14 our colleagues to design a very different
15 process for how to think about this
16 particular competition. And that process is
17 what you're participating in today.

18 We're having a number of meetings
19 around the country. We had one in Boston
20 last week. Right after Thanksgiving we'll be
21 in Denver. And we have this one here today
22 and tomorrow.

23 To get input from a number of

1 invited experts we reached out to some very
2 distinguished the boards -- the National
3 Academies of Sciences Board on Testing and
4 Assessment, the National Technical Advisory
5 Counsel, the Department of Education -- to
6 get nominations of people who they felt would
7 be wonderful experts to help guide this
8 process. And those are the people that we'll
9 be hearing from over the course of these
10 different meetings in the different cities.

11 The goal that we have is to get
12 this kind of expert input, as well as public
13 input, at the front end of the process so
14 that when we design a notice it says what we
15 think is the right thing to say and it means
16 what we mean to say. So we will not have a
17 public comment period on this. We're having
18 instead a public and expert input period at
19 the front end.

20 So we plan to release the final
21 notice inviting applications sometime in
22 March. Applications would be due in June and
23 awards have to be granted by September. This

1 is part of the Recovery Act funding we're
2 using here and all of those have to be
3 awarded by the end of September 2010.

4 So that's the process that we're
5 using here. Okay. I'm the guinea pig trying
6 to figure out where the wireless signal goes
7 in the room. There we go. And I still
8 didn't figure it out. I think the answer's
9 wave the wand around the room until the slide
10 changes.

11 Okay. So a little bit more about
12 the goals of the assessment program. You
13 probably know that we did put out a
14 preliminary notice that invited people to
15 attend these meetings and put forth their
16 both some given constraints goals that we had
17 for the program as well as asked a number of
18 questions -- and it's those questions that
19 we're going to be engaged in talking about
20 today.

21 But to give you a sense of the
22 overarching goals of the program the first
23 one is that what we wanted to do with this

1 funding was support states in delivering a
2 system of more effective and instructionally
3 useful assessments than perhaps the
4 assessments that we've been using for the
5 past several years are.

6 So the first thing is more
7 accurate information about what students know
8 and can do, achievement against standards,
9 student growth, whether students are on track
10 to being college and career ready by the time
11 that they complete high school.

12 We want these assessments to
13 reflect and support good instructional
14 practice. And we want to be sure at the
15 front end that whatever assessments we design
16 include all students, including English
17 language learners and students with
18 disabilities. We don't want this to be an
19 afterthought that we try to tack onto the end
20 of the test. We want to design the test from
21 the front end with all of these students in
22 mind.

23 We also need these tests to

1 produce data at the back end that is -- that
2 are usable to inform teaching, learning, and
3 program improvement first and foremost, but
4 also determinations of school effectiveness,
5 determinations of principal and teacher
6 effectiveness, determinations of individual
7 student college and career readiness.

8 One of the things that we'll just
9 acknowledge up front is an uncomfortable
10 order in which we're doing this is that we
11 are coming to the tail end of No Child Left
12 Behind but haven't yet started the
13 reauthorization process for the Elementary
14 and Secondary Education Act.

15 We do assume that whatever
16 assessments are designed here would be
17 assessments that would carry us across that
18 transition. So we're trying to design
19 assessments that both meet the current legal
20 requirements that we have as well as
21 transition and carry us through into whatever
22 that next set of requirements will look like.

23 So as a result of that we have

1 defined in our notice that we need these
2 tests at a minimum to cover reading, language
3 arts, and mathematics in grades 3 through 8
4 and high school. Although as you'll probably
5 hear from the experts today that doesn't mean
6 necessarily that it looks like what the tests
7 look like today.

8 So we are talking again at a
9 minimum about summative assessments, but that
10 doesn't necessarily mean that they're only
11 given at the end of the year. It doesn't
12 mean that they're only given once throughout
13 the year. It doesn't necessarily mean that
14 there's one test that can do all these
15 things. There might be a suite or system of
16 tests that we think is the right answer. So
17 those are all the questions that we've asked
18 our experts to talk to us about today.

19 We are assuming that these tests
20 may replace what's happening and what's being
21 used out there. They're not necessarily --
22 they're not additive. They could be, but
23 they certainly don't have to be. And, of

1 course, the requirement we always have that
2 they are valid, reliable, and, in particular,
3 fair to students.

4 So those are sort of the high
5 level goals. You can look in the notice that
6 we posted if you'd like to see the longer
7 list of requirements and characteristics that
8 we posted there. But those are the most
9 important framing ones for the purposes of
10 this meeting.

11 And then let me just spend a
12 couple of minutes talking to you about the
13 goals we had for this meetings. Our hope in
14 having these meetings was first of all that
15 we might be able to paint a vision of what
16 the next generation of assessments could and
17 should look like.

18 One of the things that I think is
19 so hard for us today as we're having these
20 conversations is that we all have a very
21 clear vision of multiple choice bubble tests
22 and we don't really have a very clear vision
23 of what else assessments could look like,

1 should look like, if they met all the goals
2 that we just talked about, what would you do.

3 And in any notice that we at the Department
4 put out it's very hard to communicate a
5 vision. It's going to end up looking like a
6 bunch of regulations.

7 And so we thought and hoped that
8 conversations like this that we might have
9 around the country would start painting a
10 picture for the states and the for the
11 Department of what we could and should be
12 shooting for in a way that was richer and
13 more descriptive than we'll probably be able
14 to do at least in the writing in the
15 regulations themselves.

16 But at the same time we don't just
17 want this big picture 35,000 foot view. We
18 really need very specific, very concrete
19 input and guidance from the experts we've
20 invited. So we've asked them essentially to
21 put themselves in our shoes and say, if you
22 were writing this notice what would you
23 recommend? What do you think is the right

1 thing to do for assessments? And those are
2 the questions that we'll be hearing about in
3 a minute.

4 And the other reason that we
5 really are so happy that so many states are
6 coming to these meetings that in the end the
7 states are the ones who have to come up with
8 the proposals for how to do this. And we
9 wanted to make sure that we were providing
10 through this vehicle every opportunity
11 possible for the states to hear what we were
12 hearing, to learn from what we were learning,
13 and to be able to ask questions and benefit
14 from all of the advice that we're being given
15 here.

16 So the agenda for the day looks
17 like this. We're coming to the end of the
18 stage setting section. And from here we go
19 into presentations from each of our six
20 experts -- we'll do introductions in a
21 minute -- where they will give presentations
22 and answer to the questions we asked in the
23 notice for the first 20 minutes and then

1 about 10 minutes of clarifying questions.

2 And you'll hear all of us up here
3 asking questions. We're trying to make this
4 almost like a fish bowl conversation as
5 opposed to formal presentations so that we
6 can get richer and deeper answers to the
7 questions that we might all have. In a
8 minute I'll tell you how to get your
9 questions into the mix.

10 So we do four of those before
11 lunch, two of them afterwards, and then we
12 have a roundtable -- an extended roundtable
13 discussion with all of our experts. The last
14 hour of the day is for public speakers.
15 There are a number of people who've signed up
16 speak. They'll have five minutes each to
17 speak at the very end of the day in order.
18 So that will be the concluding part of the
19 day. And that's what it looks like.

20 A couple of housekeeping things.
21 The first one is, I think when you all
22 registered you received a bunch of index
23 cards. You are welcome as we go throughout

1 the morning to write down any questions that
2 you have on the index cards. They can be
3 questions that were spurred by things people
4 up here said or questions spurred by the fact
5 that people didn't say something that you're
6 interested in and think we need to be talking
7 about.

8 We'll be collecting those from
9 you. You can always run them out to the
10 registration desk. They'll be sneaking up
11 here and bringing us the questions. And when
12 we get to the roundtable discussion we'll be
13 feeding your questions into that
14 conversation. If you have particular
15 clarifying questions for an expert just --
16 let's see, is Anya in here? So, Mark, raise
17 your hand. So Mark will be able to get
18 questions. And there's Anya at the back --
19 wave.

20 So feel free to get your questions
21 to either of them and they'll run them up
22 here to us. So if there are questions that
23 are clarifying questions for experts we get

1 those in a timely fashion.

2 We do have all kinds of cool
3 little time keeping devices up here to keep
4 us on schedule, but you might notice me
5 cutting in and being a little harsh with this
6 because we do want to make sure that we have
7 time for everyone's point of view to be heard
8 and represented.

9 I think all of you know that
10 today's section is being transcribed. We're
11 posting -- we plan to post the transcription,
12 all the PowerPoints you see, and any other
13 written input that we receive from the
14 experts or from the public on our web site.
15 There's the web site address --
16 racetothetop -- oh, that's how to submit
17 information --
18 racetothetop.assessmentinput@ed.gov. And in
19 a minute we'll show you what -- or where you
20 can go to find the information that's up
21 there.

22 I want to thank the states who
23 have come today, many from a long distance in

1 order to listen and participate in these
2 meetings today. I hope -- I sincerely hope
3 that this is -- turns out to be a good use of
4 your time. We'll certainly do our best to
5 make it that. And please do feel free to
6 submit questions and let us know if there's
7 things that you need us up here to be talking
8 about and addressing in order to help you
9 think through what your proposals might look
10 like.

11 Let me spend a minute by just
12 going around the table and letting each
13 person introduce themselves to you. Let's
14 start with you, Tom.

15 MR. FISHER: I'm Tom Fisher. I
16 most recently was the director of assessment
17 testing services in the state of Florida.
18 I'm retired.

19 MS. SISKIND: I'm Terry Siskind.
20 I'm the deputy superintendent for the
21 Division of Accountability in South Carolina,
22 and at some parts will be representing South
23 Carolina and at some parts not.

1 MR. NICHOLS: I'm Paul Nichols.
2 I'm vice president of Psychometric and
3 Research Services for Pearson.

4 MS. CAMPBELL: Good morning. I'm
5 Carol Campbell. I'm chief research officer
6 for the Ontario Ministry of Education in
7 Canada, and I previously worked in England in
8 both government and academic roles involving
9 assessment. I'm originally from Scotland --
10 that's where the accent's from.

11 MR. FABRIZIO: I'm Lou Fabrizio,
12 and I could use my growing-up accent, but I
13 won't. I'm the director of accountability
14 policy and communications with the North
15 Carolina Department of Public Instruction and
16 also a member of the National Assessment
17 Governing Board.

18 MR. GONG: I'm Brian Gong, the
19 executive director of the Center for
20 Assessment, which is a small non-profit
21 consulting firm that works with states on
22 large scale assessment and accountability
23 issues.

1 MS. WEISS: I'm Joanne Weiss,
2 director of Race to the Top Fund.

3 MS. POSNY: And I'm Alexa Posny.
4 I'm the assistant secretary for the Office of
5 Special Education and Rehabilitative
6 Services.

7 MS. WURTZEL: I'm Judy Wurtzel.
8 I'm the deputy assistant secretary for
9 planning, evaluation, and policy development
10 at the Department of Education.

11 MS. WHALEN: And I'm Ann Whalen
12 with the Office of the Secretary and
13 Department of Education.

14 MS. WEISS: Okay. With that -- I
15 don't know about this thing -- this wireless
16 stuff. There we go.

17 So this is the web site. It's
18 also published in the notice if you want to
19 find any of the materials
20 ed.gov/programs/racetothetop-assessment. And
21 that's where you can find all of the written
22 input transcripts and so on.

23 And with that I think we are ready

1 for Terry Siskind's presentation. And you
2 can pass this down. I bequeath you the magic
3 wand, Terry. Maybe you can figure out how it
4 works better than I did. Or maybe you could
5 just hear my presentation all over again.

6 (Pause.)

7 MS. WEISS: Oh, no. This happened
8 to us at the technology panel too I want you
9 to know. And we're talking PowerPoint. It's
10 not like we're talking about some fancy
11 technology, so give us a moment.

12 Okay. So let's do Plan B, Mark.
13 Let's have Lou go first and we'll have Terry
14 go after lunch.

15 MR. FABRIZIO: There goes my
16 planning time. All right. Good morning
17 everyone. Thank you, Brian. Now who do I
18 look at, Joanne? Do I look at us or talk to
19 everybody.

20 MS. WEISS: Whatever you're
21 comfortable doing.

22 MR. FABRIZIO: Okay. All right.
23 I'm going to start out with a big disclaimer,

1 and that is that anything I say that's
2 absolutely stupid or horrifying is no
3 reflection of the North Carolina Department
4 of Public Instruction, the North Carolina
5 State Board of Education, or the National
6 Assessment Governing Board. Anything really
7 brilliant I say I attribute to all three of
8 those groups.

9 Now, Joanne already set the stage
10 for the goals of the presentation. And I
11 just wanted to list some of them because as I
12 was going through the regulations all I kept
13 saying was, Wow, wow. I mean, all of these
14 things sound wonderful -- internationally
15 benchmark, college/career ready, on track for
16 college and career, accessible.

17 But when I look at all of those
18 things I don't necessarily feel like there's
19 a heck of a lot of flexibility if this is
20 going to become a checklist of what needs to
21 occur. And so I guess my first comment was
22 this is a very tall order for any of us to
23 solve and all agree with all of these

1 different parameters. And I'm not saying
2 that any of these parameters are not good --
3 or characteristics or not good
4 characteristics, but I think we could talk
5 for 20 minutes on each single one of these
6 bullets.

7 And I guess I'd like to start in
8 the beginning. And, for me, I'm hoping that
9 we get all of the sequences down correctly
10 because I know right now we're all working
11 toward the common core standards and states
12 then have to approve them or agree with them.

13 And I think it's going to be a little bit
14 complicated, especially with that 85 percent
15 rule, for all the states to be on board and
16 then for us to be able to start developing
17 those assessments as quickly as possible.

18 But irrespective of that it seems
19 like the sequence should be develop content
20 standards, then we need to make sure that
21 there is a heck of a lot of professional
22 development so teachers now know what it is
23 they need to be doing. And then we typically

1 run into the problem of we do field testing
2 of items on material or subject content that
3 teachers may not have taught the students and
4 we get these horrible item stats and we try
5 putting tests together.

6 And then, lo and behold, when we
7 find out that teachers teach the contents
8 that the kids actually can answer most of
9 those questions. And so I'm afraid that if
10 we start doing things out of order it's going
11 to cause some problems.

12 And then one of the things that we
13 were told is not to talk about
14 accountability, but I have a hard time
15 differentiating -- and I know I can separate
16 them in my head assessment and
17 accountability. But I think some of these
18 general characteristics to me smack very
19 highly of accountability.

20 So here was my, you know, earth
21 shattering recommendation to the general
22 questions one and two. One is that it would
23 be ideal for us as groups of states to

1 develop a specific set of interim assessments
2 that would be administered periodically
3 throughout the school year specifically for
4 diagnostic information. Okay? That's really
5 clear. The purpose of those interim
6 assessments is for that diagnostic
7 information.

8 Now, I'm going to break the rule
9 because my end of the year assessment, which
10 I'm saying for summative, is not for
11 diagnostic information. I know that I'm
12 violating one of the required
13 characteristics, but if I'm asked to present
14 what I think should be the case then that's
15 what I think I should be.

16 Now, I am going to throw in the
17 phrase formative assessments, but I think of
18 formative assessments as prepackaged sets of
19 questions, but it's more a process that
20 teachers use in daily instruction with
21 students. So I wanted to make sure I had all
22 of the right buzz words.

23 So, now, when I talk about those

1 interim assessments that's where I'd like to
2 see a lot of the creative stuff -- the
3 combination multiple choice, constructed
4 response, occasional performance tasks. This
5 is where it should be criterion referenced
6 because, again, we're trying to get
7 diagnostic information, and that information
8 is going to be useful for the student and the
9 teacher.

10 And then I throw in these two
11 things about how it's the interim diagnostic
12 assessments that really should become the
13 basis by which principals find out how good a
14 job a teacher is doing in changing
15 instruction based on that information.
16 Likewise, central offices need to be using
17 that information to see what a principal's
18 doing to help teachers do a better job of
19 instructing students.

20 So those last two bullets from me
21 are really more leadership roles. Now, I
22 don't remember seeing leadership anywhere in
23 these regs, but that's where I think that

1 needs to play a very important role.

2 The end of the year assessments --
3 those are going to be mostly multiple choice.

4 We probably have to throw in a few
5 constructed response items just so people
6 don't get mad about having a strictly
7 multiple choice test.

8 Ideally it would be wonderful if
9 it could be computer adaptive. And I again
10 make the comment no need for subtest scores.

11 This is just an end of the year score. It's
12 to give you a scale score. It's to tell you
13 where the kids are. It can be used for
14 monitoring growth. But it is surely not
15 going to be used for providing diagnostic
16 information to students.

17 We've got a whole 180 days of
18 instruction with students and use of interim
19 assessments to know what a student's
20 strengths and weaknesses are. And I
21 personally like the norm referenced. I like
22 for a parent to know how well their child is
23 doing compared to other students.

1 In terms of teachers scoring
2 constructed response items, this is probably
3 one of many ways it could be done. I would
4 envision the constructed response items to be
5 physically separated from the multiple choice
6 so that we could image scan the students'
7 responses. We could use distributive
8 scoring, which I know has been done in
9 several locations where we could involve
10 teachers in doing it.

11 There could be regional centers
12 that would be set up. I like the idea when
13 possible to use the automated scoring -- or
14 artificial intelligence. I don't think the
15 world is ready for that to be the only way
16 the stuff is scored, but it could be a good
17 backup reader.

18 And, again, professional
19 development and training -- I don't know how
20 many times I would want to say that for that
21 to be the key to our success in changing the
22 world here. And then exploring learning
23 management systems should become part of this

1 overall picture of what we're trying to do.

2 Purposes of assessments -- I guess
3 this is where we -- it appears to me we're
4 trying to force fit and distinguish
5 assessment from accountability. But, you
6 know, in order to develop the tests we need
7 to have the right purposes in mind.

8 And now I'm going into my random
9 other areas. We as panel members were told
10 to specifically talk about our ideal system.

11 And then there were a number of other areas
12 that we could comment on. So I'm going to
13 run through those.

14 First one I wanted to comment on
15 is the high school assessments -- and it's
16 the big distinction between end of course
17 assessments versus comprehensive exams. And
18 I don't see how you really can make decisions
19 about you're going to determine one or the
20 other and -- until you know who you're going
21 to hold accountable.

22 I mean, are you holding the
23 student accountable? Because if you hold the

1 student accountable you're going to run into
2 some issues that we've discovered in North
3 Carolina. We at one point did have a
4 comprehensive exam and we heard complaints
5 about, well, who's responsibility is it if
6 the student doesn't do well? You know, which
7 teacher can you point to? And it's great to
8 say all teachers are involved, but that
9 doesn't usually fly very well in many school
10 systems or in many school buildings.

11 Folks like the idea -- principals
12 like the idea of if I'm teaching -- testing a
13 kid on math I want to point to the math
14 teacher who's in charge of teaching that kid
15 so I can know who I'm going to hold
16 accountable if that teacher -- or if that
17 student doesn't well. So I think those are
18 questions that need to be answered.

19 And then the next question becomes
20 high stakes for students. It would be -- you
21 know, it would be nice to have some type of a
22 comprehensive exam that could be used to give
23 an indicator of how well a student is doing

1 and whether or not they are prepared. But
2 should that then become some type of high
3 stakes measure is a question that needs to be
4 answered.

5 And then when we talk about
6 college and career ready the questions that
7 always come to my mind are which colleges are
8 we talking about and which careers are we
9 talking about? Because I can assure you that
10 most kids who graduate from high school can
11 probably find some college that would accept
12 them. And so when we talk about college
13 ready and career ready I think those are just
14 very broad statements and we've got to figure
15 out what do we really mean by that, and I
16 think there's a lot that we can learn in
17 doing so.

18 Assessing students with
19 disabilities -- I'd almost rather start with
20 the last bullet first. In my opinion I still
21 don't believe that we have found the right
22 balance between doing what we believe is
23 right for the students and how we hold

1 schools accountable for how well those
2 students perform.

3 We've got all sorts of things
4 going on across this nation in terms of
5 certain students with disabilities. And --
6 maybe I'll go in reverse. And then we come
7 up with all of these accountability rules
8 that I'm not supposed to talk about, like the
9 1 percent and the 2 percent rules.

10 But all of those things drive what
11 happens in schools. And so periodically I
12 get asked to go and meet with teachers that
13 work with students for which they look at me
14 at the end of a work day and I'm shocked that
15 they can still even laugh because they start
16 describing the students that they've worked
17 with during that day.

18 And those teachers want to do
19 what's right for students. And then we come
20 up with accountability rules that make them
21 feel like their worthless as teachers. And I
22 wish we could do a better job of figuring out
23 how we can do a better job of assessing these

1 students, but, more importantly, how we hold
2 schools accountable for it.

3 Comparability of item types -- I
4 love seeing the wonderfully new technology
5 driven simulations, creative item types. I
6 think they're fantastic. But I can assure
7 you that for many of those creative item
8 types there are going to be issues that
9 hopefully we can solve that someone will
10 claim is inaccessible to some students.

11 And I just -- again, how do we
12 find this balance between those creative item
13 types that can work with many students -- how
14 do we solve that problem of those students
15 for which we may not be able to have
16 something that we can offer that one would
17 look at and say is comparable.

18 So one of the things that I think
19 is a big issue is the issue of comparability.

20 We get hung up on online tests that have to
21 be comparable in terms of whether they're
22 multiple choice on paper versus online on the
23 computer. And we spend all of this time and

1 energy doing these comparability studies, and
2 it seems like it just causes maybe more
3 problems than we're really solving because I
4 would argue that as we continue to move in
5 this area of the use of technology that the
6 kids are going to be different, what we're
7 doing is going to be different, and
8 adjustments that we may make at one point in
9 time may change later on.

10 So it's just -- it -- again, I
11 don't think we've found the right balance
12 yet. And then my first one was just supposed
13 to be funny. Technology's a wonderful thing
14 when it works.

15 In North Carolina we have over 200
16 foreign languages spoken by students in our
17 schools -- and that's in North Carolina. So
18 when I'm then asked about developing tests in
19 native languages my usual response is it's
20 hard enough for us to develop tests in
21 English. And for me to want to then tackle
22 developing tests in all of these foreign
23 languages would be a great challenge.

1 However, I know there are some
2 states that have developed tests in foreign
3 languages. And I think it would be wonderful
4 if states could purchase those tests or have
5 agreements to use those tests when we get to
6 the point where we have these comparable
7 common standards. Because if there are some
8 states that do have the wherewithal to
9 develop tests in Russian then it would be
10 great for a state to be able to use those
11 tests for those students for which it is
12 appropriate.

13 But if that's going to cause peer
14 review problems and also to other things and
15 all comparability studies that need to be
16 done then I think again we're just shooting
17 ourselves in the foot.

18 And then whole point of testing
19 students in their native language only works
20 for certain students. If the student was
21 never educated in that native language then
22 giving them a test in their native language
23 is not going to really be all that helpful.

1 So I have the same bullet on this slide as I
2 did with the students with disabilities --
3 finding that right balance between how we
4 assess these students and how we hold schools
5 accountable.

6 These are last two slides and
7 I have three-and-a-half minutes. I think
8 it's unfortunate that we're being told that
9 we want people to be creative, we want people
10 to do different things, but, oh by the way,
11 you have to keep doing what you've been doing
12 to meet NCLB requirements.

13 And it drives teachers nuts when
14 you talk to them about that kind of a system
15 because -- and we find this all the time --
16 when our state board changes its content
17 standards what do we end up having to do? We
18 end up having to say to the teachers, okay,
19 teachers, your students are going to take
20 field tests at the end of this school year,
21 and those field tests are based on the new
22 content standards that we want you to be
23 teaching this year.

1 And then hands go up in the room
2 and they say, Can you tell me how I'm going
3 to be held accountable? And we say, Oh,
4 you're going to be held accountable on the
5 old tests based on the old standards. And
6 they go, Oh, okay, so what do you want us to
7 do? And we say, Oh, nothing really hard. We
8 want you to teach the old content standards
9 so your kids will do well on the old tests,
10 but we want you to teach the new content
11 standards so that we can get really good
12 reliable field test data from your students.

13 And I would plead with the U.S.
14 Department of Education to work with the U.S.
15 Congress to figure out better ways of doing
16 this. And one way I think is you phase in
17 the subjects over time, you require statewide
18 field tests in the subject for which you're
19 going to do the field testing, and you do not
20 put those students through an operational
21 form of the official test so that you can
22 truly say to teachers, Mel, this school year
23 here is the new content you are supposed to

1 be teaching -- teach it. We're going to give
2 you professional development and training and
3 we want you to do a great job.

4 When we talk about transitioning
5 to new assessments the devil is in the
6 details. I have a special space in my office
7 for that person and we get along great. We
8 also need to make sure that we set aside the
9 necessary time to do this correctly. I
10 understand that the U.S. Department of
11 Education has this once in a lifetime
12 opportunity. I would just say be very
13 careful that we don't do things in such a
14 rushed fashion that we don't really get what
15 we're hoping for.

16 I think this endeavor is going to
17 be very expensive. I did not cost it out --
18 I can just tell you it's going to be
19 expensive just from what we know. I think we
20 need to have conversations long and hard
21 about intended versus unintended
22 consequences. And I think we really need to
23 find the right balance between instruction

1 and assessment. I'm really tired of the
2 amount of frustration that I see and witness
3 all the time about the problems with
4 assessment when the focus needs to be on
5 instruction.

6 And my final bullet -- I think any
7 kind of collaborations that can occur with
8 universities is the right thing to do,
9 especially in light of training future
10 teachers or training teachers in teacher
11 preparation programs. They need to be
12 involved in this effort. My time is up.

13 MS. WEISS: Perfect timing. Thank
14 you. Questions?

15 MS. WURTZEL: Thanks very much,
16 Lou. So I want to actually speak to one of
17 your last points, which is about the balance
18 between instruction and assessment. And
19 given what we know about the power of
20 assessment to drive instruction, can you
21 elaborate a little bit more why in your
22 proposal you suggest that the summative
23 assessment be multiple choice and constructed

1 response, which are -- the assessment that
2 has stakes that may drive instruction --
3 while it's the interims where you want more
4 performance oriented tests and how you think
5 that's going to get us to the kind of
6 instructional improvements that we hope to
7 see?

8 MR. FABRIZIO: In my opinion for
9 all of this to work we need to be having good
10 things happening in classrooms each and every
11 day. And if the teacher's usual complaint
12 from the end of the year test is that it's
13 got high stakes, it's what's going to
14 determine whether or not we've done a good
15 job as teachers.

16 I think we've got to turn that
17 back onto teachers and say, No, it's what you
18 do during the day-to-day stuff that really
19 counts. And we're going to give you some
20 periodic tools to use to determine how well
21 your students are doing.

22 If those become the high stakes
23 for teachers then we're going to have schools

1 developing practice tests for the interim
2 assessments and we will have testing again
3 driving inappropriate practice. And that's
4 why I believe if we could do a better job of
5 getting those teachers to see the benefit of
6 the constructed response items and the
7 performance tasks for what they're doing
8 during that school year that that will work.

9 If we're going to then have this
10 big push on looking at, quote, growth, I
11 think that growth can just be a single number
12 because that's usually any -- any growth
13 studies that I've seen usually come up with
14 one number that's used for multiple points
15 over the years. And I would do whatever I
16 could to minimize that particular assessment.

17 MS. POSNY: Lou, to follow up on
18 that, when you talk about the growth are you
19 talking about the growth only on the
20 summative measure and not the interim
21 measures?

22 MR. FABRIZIO: That's correct.

23 MS. POSNY: Okay. Let me ask

1 another question to go along with this
2 because I think you'll know where I'm headed.

3 If -- and, again, you talked about an
4 adaptive test that could possibly be used as
5 part of the interims, but, again, not the
6 summative.

7 But if you really think about an
8 ideal system -- and I'm going towards some
9 of -- you know, some of the special
10 populations that we're talking about -- is
11 there not a way to build the system that
12 might allow the adaptive to be part of a
13 summative measure or, you know, the interim?

14 I'm just trying to figure out how we wrestle
15 with the different needs and the different
16 students and use the system you're talking
17 about.

18 MR. FABRIZIO: Well, I think that
19 would be the beauty of with this new \$350
20 million -- it should be used for trying out
21 all these different things. But if you're
22 going to then put parameters around what all
23 those things have to be I think you're going

1 to lose the opportunity that we're trying to
2 find.

3 I'd love to have -- well,
4 here's -- everybody wants a ten-item test
5 that takes ten minutes to administer that
6 gives you ten pages of diagnostic
7 information. I mean, that's what we all
8 want. And anything that we can do that gets
9 us to that I think would be great.

10 MS. POSNY: Okay. This is going
11 to be a tougher question. What about -- in
12 the ideal world what time frame are you
13 talking about? Three? Five? Ten years? --
14 seriously, if you had the time to develop the
15 system the way you thought it needed to be
16 developed.

17 MR. FABRIZIO: I think three years
18 would be unrealistic. Five years is
19 possible. Ten years is more likely. But,
20 again, if people have to keep doing what
21 they've been doing while then trying to do
22 this experimental stuff I think that's where
23 we run into the problems we've run into.

1 MS. WEISS: Other questions?

2 MS. WHALEN: I had a quick
3 question. I was wondering if you could
4 elaborate a little more about what you're
5 thinking with these diagnostic assessments?
6 Would it be analyzing the students'
7 performance on the lesson just taught so it
8 would be a test that's developed particularly
9 to the scope and sequence of the curriculum
10 being implemented? Or would it be just a
11 kind of sampling of what will be on the
12 summative end of year assessment? And if
13 it's the former would that then imply that
14 the consortia that would apply and implement
15 these assessments have to have a similar
16 scope and sequence, or can you envision
17 different ways for that to happen?

18 MR. FABRIZIO: I think for the
19 benefit of the students it really should be
20 more on what's been taught during a specific
21 period of time. I would not be necessarily
22 advocating that that occurs once a week. But
23 when I talk about those interim assessments,

1 you know, some folks refer to them as
2 benchmark assessments.

3 From everything that I've pretty
4 much seen though when you administer specific
5 tests and then you try linking it to that end
6 of the year test what we typically find are
7 correlations somewhere in the 60, 70 percent,
8 which basically says kids who tend to do well
9 on this test tend to do well on this other
10 test.

11 And so if the main purpose is to
12 find out whether or not the teacher did a
13 good job of teaching certain things I think
14 that scope and sequence is going to have to
15 be part of it. And so if states are supposed
16 to have common core standards somebody's got
17 to figure out at least some parameters of
18 about how much time -- you know, we call them
19 pacing guides in North Carolina.

20 But I don't -- I hopefully -- do
21 not hope that we're going to get to the day
22 where, you know it's Tuesday, March 13, and
23 so everyone's on this particular activity.

1 But I do think that would make more sense to
2 have tests based on what the student --
3 rather than it being an alternate form of the
4 end of the year test that you're just giving
5 two or three times during that school year.
6 I don't see any value with that.

7 MS. WEISS: Paul, did you have a
8 question?

9 MR. NICHOLS: Yeah. So, Lou, you
10 seem to be referring to tension in the
11 consequences -- unintended consequences. Can
12 you hear me okay?

13 MS. WEISS: I'm not sure. Can you
14 just make sure the volume levels are high
15 enough?

16 MR. NICHOLS: So you've got
17 unintended attention and unintended
18 consequences. You've got a summative
19 assessment that's different in format than
20 your classroom based assessment. So that
21 could undermine teachers' use of classroom
22 based assessment because it's a different
23 format. So they're practicing on multiple

1 choice tests rather than the sort of
2 portfolio and performance based tests. On
3 the other hand, if classroom based assessment
4 counts as -- in a summative way then that's
5 undermining teachers' use of that.

6 Is there any way to resolve that
7 tension? Can you think of any way to resolve
8 it? Could you combine the two in some way?
9 I know some systems in England or Scotland,
10 for example, do that. Do you see any way to
11 combine the two to resolve that tension?

12 MR. FABRIZIO: That was one of big
13 struggles that I had. I actually had a
14 bullet on my slides that I then removed,
15 which was trying to specifically combine
16 those two types. And I had the phrase, If
17 psychometrically appropriate.

18 And then I started thinking about,
19 you know, well, what are the unintended
20 consequences that are going to occur, and
21 that's when the idea of people then coming up
22 with practice tests for the interim tests.
23 So now we've just gone from having one end of

1 the year high stakes test to six to ten
2 during ten year.

3 I'm not opposed to it. I mean, if
4 it can be demonstrated that it can work I'm
5 all in favor of it. But, again, I think
6 unless we give these different state
7 consortia more free reign to do things from
8 which we can learn new things I don't think
9 we're going to solve it.

10 MS. WEISS: Brian, we'll give the
11 last word to you and then we'll move on to
12 Karen.

13 MR. GONG: A really quick
14 question, Lou. Your ideal system is one that
15 you could pursue in North Carolina by
16 yourself. What could the U.S. Department of
17 Ed do in this common assessment RFP that
18 would help you do -- achieve that through
19 working with other states?

20 MR. FABRIZIO: It would be that
21 recommendation regarding the -- I'll call it
22 the waiver -- the waiver of having to
23 administer an operational test in the year

1 that you're doing the field testing.

2 (Pause.)

3 MR. FABRIZIO: And, Terry, you owe
4 me for going first.

5 MS. WEISS: Carol?

6 MS. CAMPBELL: Thank you. I'm
7 actually going to stand if you can still hear
8 me, because I can't see half the room if I
9 talk from here.

10 I'm going to bring a slightly
11 different perspective -- and, of course,
12 that's deliberate. I come from Canada and
13 the U.K., and I want to talk a bit about
14 international systems, particularly from
15 those two that I know well.

16 So I'm not going to get into all
17 the details that some of the other panel
18 members will, but to talk about how we look
19 at assessments and what implications that
20 could or couldn't have for your debate here
21 in the U.S.

22 Joanne asked us to paint a vision
23 but also provide concrete examples, and

1 that's what I'm going to try to do in the
2 next 20 minutes. I'm going to go fast. But
3 I started to think about what could be a
4 vision for assessment, thinking on some of
5 the practical experiences that I've been
6 involved in in the systems I've worked in,
7 but also to think about giving you some
8 examples of what these look like. So it's a
9 combination of the principles and some
10 examples.

11 And these were ten principles I
12 thought about as I reflected on assessment,
13 designed around a clear vision of purpose and
14 learning goals, and I think you're hear that
15 as we go through the presentations.

16 Align with and advance common
17 standards -- the debate that's happening --
18 curriculum expectations; learning objectives
19 and instructional strategies; encompass and
20 combine assessment all for and as learning;
21 use and blend multiple sources of information
22 and assessment approaches to identify,
23 inform, and report student learning; build

1 and develop professional capacity -- and I
2 absolutely echo what Lou said on that.

3 Embody high expectations to engage
4 more to the support and stretch all students
5 to progress and achieve; assess a range of
6 content, knowledge, skills, and performance
7 in authentic, applied, and appropriate
8 ways -- and that is complex and challenging;
9 think about assessment as part of a clear and
10 flexible learning trajectory across subjects
11 and grades but also this to be around career
12 and work and education.

13 In thinking about the systems that
14 need to be put in place alongside assessment
15 around data informed decisions, classrooms,
16 school districts, state, and indeed federal
17 level; and, finally, place quality over
18 quantity.

19 So I'm going to give you some
20 examples to illustrate what I mean. On the
21 first one, thinking about clear vision of
22 purpose and learning goals -- Lou's already
23 said which careers, which colleges, what

1 future: the whole debate of a 21st century
2 skills.

3 The example I'm going to use is
4 from England, where they're looking very
5 carefully at both subject knowledge but also
6 a whole generic set of skills that could be
7 called 21st century skills across all the
8 curricular areas.

9 So as well as the detailed
10 curriculum there is what's called functional
11 skills, and these are designed as essential
12 for life, learning, and work. And they're
13 very practical skills in English,
14 mathematics, and information and
15 communication technology. They're applied
16 skills, and the actual examples are from work
17 experience and work examples.

18 Then a whole set of personal
19 learning and thinking skills: PLTS. These
20 are skills that will equip people for
21 successful employment and lifelong learning.

22 And the ones that are defined by the
23 government in England are independent

1 inquirers -- learners as independent
2 inquirers, creative thinkers, reflective
3 learners, team workers, self-managers, and
4 effective participants: the types of
5 qualities that we need in the workforce, but
6 also in future education.

7 And you can say, All well and
8 good, how the heck do we measure those? So
9 some of the work that's been done in pilots
10 over the past three years is actually looking
11 at criteria around some of this sort of
12 skilled development and types of knowledge.

13 And these are some of the criteria
14 that the assessment bodies in England have
15 developed. For functional skills or
16 practical skills, the skill areas must, the
17 assessments must provide realistic context
18 scenarios and problems; specify tasks that
19 are relevant; require application of
20 knowledge, skills and understanding; require
21 problem solving; assess process skills,
22 things like team work and communication in
23 different contexts.

1 And these combine a range of
2 external and internal assessments. So there
3 is external criteria from exam boards around
4 what are the requirements that support the
5 nature of the assessments. But there's
6 course work and project work and teacher
7 involvement in the actual assessment process
8 linked to the formative debate.

9 And what's called controlled
10 assessments -- so this is assessment within
11 the classroom context, course work, and
12 homework, but it's supervised. And looking
13 at how these involve both open types of
14 questions, projects, open-ended response, but
15 there can also be some fixed responses in
16 specific subject areas.

17 This whole idea of performance
18 assessment -- actually applying your
19 knowledge and demonstrating that -- is
20 increasingly used for age 14 to 19 learners
21 in England. And there's also opportunities
22 for some diploma qualifications making that
23 link into college, where students who are in

1 schools can also have accreditation for work
2 experience.

3 This is an example from Ontario,
4 where in Ontario we have key skills across
5 our curriculum -- and I'll talk a bit more
6 about that. But there's examples that
7 supports the teachers to look at how you can
8 assess those key skills in day-to-day
9 classroom practice that then feeds into
10 annual report cards and, indeed, to our
11 provincial standardized assessment. So this
12 is an example of thinking skills -- critical
13 thinking skills in English, grades 9 to 12:
14 use of planning skills, processing skills,
15 creative and critical thinking skills, and
16 there's details underneath this I can't get
17 into. But it's looking at what specifically
18 would you look for in student work.

19 And that leads into the principle
20 number two about alignment. Align the common
21 standards, curriculum expectations, learning
22 objectives, and instructional strategies and
23 assessment, and I think that links to what

1 Lou said.

2 So in Ontario these are very
3 carefully aligned. There's a curriculum and
4 there's content standards, grades 11 to 12,
5 for all subjects and all courses. They're
6 benchmarked. All teachers are required to
7 teach the curriculum.

8 There are performance standards or
9 criterion referenced to the curriculum.
10 There's criteria described for student
11 achievement; there's four levels of
12 achievement, level one, two, three, and four.

13 There categories of knowledge and skills,
14 and all of this is demonstrated through
15 curriculum documents, materials, rubrics
16 resources, and lots of professional
17 development, which links into the
18 consistency, providing exemplars of student
19 work, of classroom practice, lessons plans,
20 and also leading into provincial report
21 cards -- reports to students, to parents --
22 that we also use for some of our assessment
23 approaches.

1 So this whole piece of assessment
2 all for and as learning, the blending of
3 while recognizing the distinctions of
4 summative assessments, tests at the end of a
5 grade year -- you know, of study -- for
6 learning the formative assessments throughout
7 the classroom process with a focus on
8 feedback.

9 And increasingly we're looking at
10 assessment as learning: how to engage
11 students themselves, self-evaluation, self-
12 assessment; that whole meta-cognitive of
13 process of knowing where you are and knowing
14 where to go next.

15 And this is an example from
16 Scotland. As Paul mentioned, where -- if you
17 think about an assessment system, there's a
18 whole range of different types of assessment
19 and they're used for different purposes. So
20 in this one there's a combination of internal
21 teacher-level classroom assessments,
22 external -- there are standardized tests, and
23 these are in the bottom quadrant: the

1 Scottish Survey of Achievement and National
2 Qualifications.

3 These are the summative tests.
4 There's a range of formative pieces along the
5 way and some that are highlight -- things
6 like personal learning plans. These are
7 plans that are developed with students --
8 teachers' judgment -- so the whole piece
9 around how we bring together classroom
10 assessment with more standardized assessment.

11 Principle number four: use and
12 blend multiple sources of information and
13 assessment approaches to identify, inform,
14 and report on student learning -- multi-
15 faceted assessment approaches using a range
16 of information.

17 Here I'm comparing England and
18 Ontario, just to give you a sense of the
19 types of assessments that actually are being
20 combined so they're proportionate. So each
21 assessment contributes a certain amount to
22 the final mark for a student.

23 So the standardized tests --

1 multiple choice response -- that's just under
2 half of the items in our standardized tests,
3 grade 9 and 10; it's about 46 percent, 48
4 percent depending on the subject area.
5 Multiple choice responses are no longer used
6 in England. They decided to move away from
7 that process completely.

8 So in England it's open-ended
9 responses in all the standard assessments,
10 and they account for up to 40 to 100 percent
11 of the final mark, depending on the subject,
12 and they're half of the score in Ontario.

13 So what's England doing? Well, a
14 whole move to performance assessment,
15 controlled assessments, course work,
16 projects; these are up to 60 percent of a
17 student's grades at age 15 and age 16. And
18 then advanced qualifications projects -- the
19 type of work that you do in college and
20 university and beyond -- thinking about real
21 inquiry projects.

22 The more systematic use of teacher
23 assessments -- gathering together classroom

1 data, report card data, and using that data
2 not just in school level discussions but how
3 we can use it at the system level -- thinking
4 around reporting and timely report card type
5 marks, working those sorts of pieces. And
6 actually in England they've abolished some of
7 the mandatory testing, and they're replacing
8 it with using the teacher assessments.

9 And, finally, the use of
10 portfolios and cumulative assessments,
11 gathering today a body of student work over
12 time.

13 I'm going to skip the next two;
14 they're in the pack.

15 So all of this is about high
16 expectations to engage, motivate, support,
17 and stretch all students, all learners with
18 all their varieties and diversity and needs
19 to both progress and achieve. It's no longer
20 just about the test result; it's about
21 student growth, to use that word.

22 And one of the examples that I'm
23 going to use is from England: this idea of a

1 making-good-progress pilot where they're
2 setting progressive targets. So looking at a
3 student's classroom work and seeing where are
4 they at the start of the year and how far are
5 they progressing in their learning over that
6 year and two years and three years. And
7 progression premiums for schools -- thinking
8 about students who enter a grade level behind
9 expectations and whether they have caught or,
10 indeed, exceeded by the end of that period of
11 time.

12 The other piece around high
13 expectations has been very clear around how
14 assessment can support high expectations.
15 And this is from the Assessment for Learning
16 Strategy in England. I won't read it all,
17 but I will flag the first piece.

18 Every child knows how they're
19 doing and understand what they need to do to
20 improve and how to get there. They get
21 support. They need to be motivated,
22 independent learners on an ambitious
23 trajectory of improvement. That's a bold,

1 big statement. But you see how assessment is
2 seen as part of high expectations for
3 teaching and learning.

4 So high expectations feed in to
5 higher-order thinking skills, critical
6 skills, 21st century skills, problem solving.

7 What does this look like in tests? So this
8 is an example from grade ten literacy task in
9 Ontario. This is our sort of graduation type
10 test:

11 Write a minimum of three
12 paragraphs expressing an opinion on the topic
13 below. Develop your idea with supporting
14 details: Are today's famous people good role
15 models for young people? And you can see the
16 whole set of critical thinking and problem
17 solving that goes into that.

18 Another example is through
19 bringing together content and knowledge
20 skills and performance and how you blend
21 those in performance assessments, as well as
22 standard assessments. And this is the
23 direction that the GCSE in England has gone

1 down very heavily at the moment. And I'm
2 going to give you an example from an actual
3 assessment there.

4 This is for information
5 technology; I'll just pull it out. This is a
6 project that a student would be involved in.

7 It would be about 30 hours of work, and it
8 would be 60 percent of their final mark. So
9 there would also be a summative test, but 60
10 percent of the mark is on project work.

11 This is a fictitious IT company.
12 There's a promotions company. They need
13 support with a database and you, as the
14 student, need to work with others to plan and
15 carry out research to investigate how similar
16 companies have produced a solution. You need
17 to record and display your findings. You
18 need to recommend a solution.

19 You need to produce a design
20 brief -- and so it goes on on the next one.
21 You need to actually produce the solutions.
22 So you need to demonstrate your subject
23 knowledge and your technical skills in IT.

1 You need to incorporate software, but then
2 you get into the user feedback, require --
3 present information, communication, and
4 interpersonal skills and evaluate your own
5 and other work. So you see that whole
6 blending of subject knowledge with functional
7 skills and professional learning skills.

8 Similarly, at the A levels, which
9 is the advanced qualifications in England,
10 the introduction of more and more extended
11 writing and what they're calling synoptic
12 assessments, students having to demonstrate
13 their knowledge across a subject area and
14 introducing project type work.

15 But critical -- absolutely
16 critical to this is professional capacity,
17 and I completely agree with the points that
18 have been made. Taking the time to build and
19 develop professional capacity and putting
20 resources and support in -- one example is in
21 Canada -- Alberta -- and I'll talk about
22 Ontario also -- teachers are involved
23 throughout the whole assessment process.

1 There's teachers involved in the test design,
2 there's teachers involved in the item
3 development, there's teachers involved in
4 reviewing the actual tests once they've been
5 constructed looking at whether they're fair
6 and appropriate, connect with students.

7 Teachers look at the standards.
8 They look at student work itself against the
9 standards and the tests. Teachers have
10 opportunities to be involved in the marking
11 process, so there is marking of open items
12 and project work directly involving teachers.

13 And there's other ways to be involved in
14 committee type work.

15 Teachers that are involved in this
16 generally say it's the most powerful
17 professional development experience that they
18 have. And while it takes time and it takes
19 resources it flows through everything that
20 happens throughout the school year.

21 We have a similar process on
22 Ontario where teachers are involved in
23 scoring the majority of open ended responses,

1 development of rubrics and scoring guides,
2 development of anchor charts -- a whole set
3 of systems and processes that I certainly can
4 talk about.

5 And this feeds into the whole idea
6 of assessment for learning. As teachers
7 become more and more and more involved in the
8 assessment process how we can also use that
9 assessment data, summative and formative, in
10 our classroom instruction. And one of the
11 models that's been powerful for us is a
12 teaching learning critical pathway -- there's
13 different names for this.

14 But basically using the assessment
15 data from end of year tests, but also
16 classroom tests and saying, Where's the areas
17 of greatest need in the curriculum and for
18 our learners -- our students -- this year.
19 What's our current practice? What's our
20 current instructional practice and how well
21 is it or not working for our students?

22 Developing data will also remap
23 our students' progress over the course of a

1 year. I don't find what instructional
2 interventions to put in place. Working as a
3 teaching, learning community, moderated
4 marking, teachers coming together with
5 student work and really debating the levels
6 and the standards and the quality, moving
7 that forward. And in our schools that are
8 using this, we're seeing double0digit
9 percentage point improvement results.

10 So onto the learner. We think
11 about assessment as being part of a flexible
12 learning trajectory. Think here about how
13 assessment and core subjects and other areas
14 supports each other. And the governance
15 around that -- I'm going to use England as an
16 example.

17 England has five qualifications
18 boards. It's not one; there are multiple
19 boards and there are differences. But there
20 is consistency of standards within a subject
21 area and flexibility across the subject
22 areas.

23 There's also a clear continuum for

1 the learner -- her trajectory -- college and
2 career ready. GCSEs and A levels are the
3 sort of traditional academic, but they're
4 becoming more work oriented as well. The
5 diploma includes vocational education, and
6 apprentices are work based and college based.

7

8 But learners can move between
9 these and they choose. They choose subjects.

10 There's no mandatory assessments. They have
11 a blend and balance between those pieces.
12 But you can see the trajectory of learning
13 there.

14 The next piece I'm not going to go
15 into in detail; when you have the PowerPoint,
16 you can see it. But the whole piece around
17 the data-informed decision making. As you
18 develop the assessments I would encourage you
19 to think about the technology and how you
20 support that in the classroom, school,
21 district, and state level.

22 And some of the work that's on the
23 slides is taking the principles of classroom

1 and school assessment, but thinking how they
2 can apply at the system level: pieces around
3 planning, how to support students in schools,
4 thinking around professional development
5 needs, thinking about a selection of types of
6 data, how you engage motivation and so on.
7 The most important piece is working to engage
8 achievement data, but that it's reflective of
9 progress, growth, and all learners' needs.

10 This is our data system in
11 Ontario. I'm not going to talk in the
12 details of that one just now.

13 So thinking about quality over
14 quantity -- I think one of the messages I do
15 want to see is that the idea of every
16 students being summatively tested every year,
17 every grade is rare internationally. It
18 tends to be that the highly standardized
19 tests are only in certain grade levels and
20 there's a lot of use of other types of
21 formative and classroom assessments built in.

22 There's a whole set of pieces
23 around that linkage between assessment and

1 instruction and standards. So I think going
2 back to the principles I'm going to quickly
3 see; start with the end of mind. I think Lou
4 and others will say the same.

5 What you want for your high school
6 students, your graduates, and the future
7 citizens of the U.S.: align as far as
8 possible assessment and curriculum and
9 instruction, development an assessment
10 system, think not only of summative
11 assessments but the whole blend of range of
12 other types of systems. Think about the
13 range of items and types of tests and other
14 forms of information that can be used.

15 Think about educator capacity --
16 absolutely critical. Think about your
17 repertoire. Think about the clarity of
18 learning pathways and how you could support
19 that. Think about the data infrastructure
20 but also the culture -- the human capacity
21 and values around these pieces.

22 And think about shifting the
23 emphasis from quantity of multiple choice

1 assessments to a whole range of forms of high
2 quality robust assessments that are complex
3 and challenging but educationally very
4 important.

5 MS. WEISS: Thanks, Carol. That
6 was terrific. That was a lot of stuff going
7 by quickly so we might have questions for you
8 that take you back to different parts.

9 Let me start by just asking you a
10 quick question about something you said about
11 something that was producing double digit
12 gains and improvement for students in Ontario
13 and ask you maybe a slightly broader question
14 about just the big picture experiences with
15 any relationships that you saw here or in
16 England between these types of assessments,
17 gains in student growth, plateaus you might
18 have seen at different points. Just how do
19 these two things seems to correlate?

20 MS. CAMPBELL: To start with
21 Ontario, the piece that I went through -- I
22 know I went through the whole thing
23 quickly -- but the piece around teacher

1 moderation, teacher engagement, and looking
2 at the student work in involving classroom
3 assessment. But also we do use our
4 standardized assessments and go back to the
5 item level detail and look at the areas of
6 the curriculum that appear to have been
7 weaker and where students aren't performing.

8 But taking all of that detail and
9 getting into very rich professional
10 conversation about where appear to be the
11 weaknesses -- at the system level
12 provincially, district level, school level,
13 but particularly the classroom level -- and
14 using that information to target
15 interventions and innovations for students,
16 but also for classrooms and for schools.

17 And through that activity where
18 the teachers come together and then we put in
19 place supports, that's the piece that we have
20 seen very substantial gains in professional
21 learning and student -- and teacher
22 understanding of the assessments, but in
23 student growth and student participation.

1 So we don't have very many very
2 low performing schools anymore, but we do
3 have schools that are plateauing, and what we
4 see there is when we get much deeper into
5 understanding the linkage between the
6 assessment and the curriculum and the
7 instruction, that's where they move again,
8 and that's because all of those pieces are
9 aligned.

10 In England the story is slightly
11 different. In England, as many of you know,
12 there's a long period of substantial focus in
13 education now, the introduction of
14 accountability and standardized assessments
15 in literacy and numeracy particularly.

16 And in the late nineties there was
17 quite a substantial improvement. So to some
18 extent the standardized tests did drive
19 improvement because they drove changes in
20 classroom practice and teaching. That
21 plateaued. That lasted about three or four
22 years. And there what we discovered was that
23 to get much deeper was to get into formative

1 assessments and teachers' professional
2 judgment around some of the classroom
3 practice.

4 So, you know, I do think one of
5 the messages is that summative assessments
6 have their purpose and their place. They
7 only take you so far when you're trying to
8 get to the higher level skills of thinking
9 and critical thinking and problem solving.
10 And we've had to look at the balance of those
11 two pieces.

12 MS. WEISS: Questions?

13 MS. POSNY: If I could ask a
14 question. It sounds like a great system.
15 You had in there common practices across the
16 countries -- not the same approach for every
17 student, every grade, every year. Is that in
18 relationship to like the interim assessments
19 or the formative rather than the summative?
20 And then my specific question is is what
21 about students with disabilities? Are they
22 held to the exact same standards and do they
23 take part in the same assessments and are

1 schools held accountable to moving up to
2 those two progress levels?

3 MS. CAMPBELL: So the -- on the
4 piece and not the same for every student,
5 every grade, every year, that applies to
6 formative and summative. And in both Ontario
7 and England in some grade levels there no
8 formalized summarized assessment in terms of
9 an external test. There would be a reporting
10 of teacher judgment based on course work and
11 classroom work and other types of tests. But
12 it's not an expectation at every grade level.

13 There is that end of year test in the same
14 way. It varies considerably by subject and
15 it varies by grade level.

16 On students with disabilities, the
17 expectation is that they're included -- as
18 fully integrated as is possible. But there
19 are, of course, opportunities for
20 modifications depending on the nature of the
21 disability and when needed. So there's
22 individual education plans and, as you know,
23 the whole range of modifications.

1 But, of course, if you're thinking
2 around a balance between course work and
3 tests there's also a whole range of ways of
4 engaging students throughout the whole
5 process. So it's not that once a time
6 opportunity.

7 MS. WURTZEL: So you've described
8 a couple of systems that used a lot of
9 locally generated assessment, data, and
10 teacher scored assessment data in
11 determinations around school effectiveness.
12 Could you give us some thoughts about
13 processes for auditing and ensuring
14 comparability and fairness in those
15 processes?

16 MS. CAMPBELL: Yeah. So on the
17 teacher assessment on the standardized
18 assessments -- the summative pieces -- it --
19 there is a rigorous process. And, you know,
20 that's something that would certainly take
21 time to develop and to work through.

22 So it's not all teachers that are
23 involved, but teachers have the opportunity

1 to be involved in marking some of the
2 standardized tests. They come together,
3 there's training, there's intensive training,
4 there's core marking and core assessments.
5 So everything is marked by more than one
6 teacher and then -- or educator. Then there
7 would be a discussion about the marks and the
8 distribution of the marks and why did you
9 think this is a level three; I thought it was
10 a level two, and you begin to see the range
11 of how many people are coming up with the
12 same level grade and what the range of them
13 is, and there's a debate about that.

14 There are expert examiners who
15 would then be involved in the process where
16 there clearly was a very broad range of
17 views. And there would be adjudication
18 process, but there's also a rubric, so
19 there's written documentation around standard
20 levels and anchor charts, the markers the
21 educators use.

22 So you have the resources, you
23 have the professional discussion, but there's

1 also a sort of system of checks and balances
2 with the senior examiners and actually there
3 comes a use of technology, where some of this
4 is scanned into computers and there's interim
5 checking to see where's the testing taking
6 place.

7 So it is quite a complex process
8 with a whole sets of layers of checks built
9 in from the marking itself through to the
10 senior examiner to the higher level exam
11 board's quality, assure and quality check.
12 Not every single test scrip, but a
13 substantial number of them.

14 On the course work piece, there
15 are exam boards in England who sets criteria,
16 and there is a regulator who regulates
17 quality. And, again, some of that's on a
18 sampling basis. But the course work is
19 either set by external criteria and marked by
20 teachers or can be locally generated but then
21 is externally marked. So it's not all or
22 one. There's that balance between the
23 external and the internal.

1 In Ontario when our report cards,
2 which is the teacher level data -- we're
3 actually at the moment of running some of the
4 correlations to see the balance between
5 formative and summative and to what extent
6 they're telling us the same or different
7 pieces.

8 MS. WEISS: Any other questions?
9 Go ahead, Brian.

10 MR. GONG: Carol, you know that in
11 the United States there is a national or
12 federal level of government, state, district,
13 school boards and so on. Could you talk from
14 your experience what the different levels of
15 governance play in the system and whether --
16 you know, there are central things that have
17 to be done at one of those levels.

18 MS. CAMPBELL: I will have to
19 think about that one. So in Canada it's
20 federal, provincial, and district. Federal
21 government does not have a direct role in
22 education so it doesn't completely transfer.
23 The provincial level, which would be

1 equivalent to the state, the -- we have a
2 strong role in assessment and in all areas of
3 education.

4 So we have a independent
5 assessment agency that we work with that
6 actually prepare the tests and develop them.

7 But the whole piece around making sure
8 there's an assessment infrastructure, there's
9 assessment policies and guidelines, there's
10 obviously the curriculum itself, there's a
11 huge investment in teacher development and
12 capacity. And we've put in substantial
13 resources into the data infrastructure.

14 So the state -- or in our case the
15 province -- has a huge role in setting sort
16 of overall guidance, frameworks, policy
17 guidance, and putting in resources to build
18 the capacity.

19 The districts do have flexibility
20 so at the district level there could -- there
21 will be variances in specifics around the
22 data system or those types of pieces, but
23 it's within the overall framework.

1 England it's nationally
2 determined, and basically the districts and
3 the schools work within that national
4 framework. But England's a smaller country,
5 of course, than the U.S.

6 MS. WEISS: Any other questions?

7 (Pause.)

8 MS. WEISS: Okay. Let's get --
9 we're going to turn back to Teri, whose
10 PowerPoint presentation has been located.

11 MS. SISKIND: Okay. I'm Teri
12 Siskind. I'm from South Carolina. I am
13 currently a deputy superintendent for
14 accountability, but I am a former assessment
15 director.

16 And when I was asked to present as
17 part of these panels, I had a dream a couple
18 of days later, and I just wanted to mention
19 this dream to all of you. The dream was
20 about Paul Sandover, who is a former
21 assessment director in South Carolina as
22 well.

23 And in this dream Paul was in a

1 car, and when he would push the accelerator,
2 his foot would go through the floorboard.
3 And when he would push the brake, his foot
4 would go through the floorboard. So I don't
5 know how many of you are into dream analysis,
6 but I thought that was something for me at
7 least to consider.

8 Secondly, I have a six-word title
9 to my PowerPoint presentation, and that's
10 purposeful. I think Hemingway said any story
11 could be told in six words; maybe it was the
12 *Onion*; I don't know. But, anyway, so I've
13 got six words, six slides, and that's where I
14 want to go from there.

15 This Venn diagram represents at
16 the top the Teri Siskind grandiose dream,
17 which I want to talk about for a minute, the
18 South Carolina Education Accountability Act,
19 and the Race to the Top. And as you can see
20 there's a very small commonality there. If
21 you add in things like ESCA, other
22 requirements -- maybe the common core --
23 there might even be a smaller area of

1 commonality there.

2 But when I first read -- and I
3 tend to, you know, not read very carefully at
4 first -- when I first read that I could have
5 a grandiose dream that's what I started out
6 with. And so that's the top -- the first
7 thing I wanted to talk about.

8 I guess in my grandiose dream what
9 I would envision is that students would have
10 a little lap book or some sort of computing
11 device, which are very reasonable now, and
12 that these devices would serve them in the
13 classroom. They would have their textbooks
14 on them, they would have other resource
15 materials on them, they would have the
16 accessibility to the internet available to
17 students.

18 And that at some point in the year
19 students would be randomly thrown assessments
20 in some -- maybe three -- of a myriad of
21 areas. So it wouldn't be limited to English,
22 language arts or reading and writing and
23 math. It would be science and social studies

1 and financial literacy and arts and P.E. and
2 foreign languages and just -- so I might
3 randomly get a writing, a Spanish, and a
4 financial literacy test one year. And that
5 this would be typical performance and that
6 these results would be reported at the state
7 level only. So the stakes would be at the
8 state level.

9 Now, I understand that due to the
10 requirements here that might have to be
11 modified so that every student was receiving
12 something in English, language arts, or
13 reading, and mathematics. But, ideally,
14 that's the kind of thing I see.

15 I also understand that it's
16 probably not practical for the purposes --
17 for your purposes to report -- or even for
18 ours -- to report only at the state level for
19 state comparisons. So I can see that this
20 might be something that would have to be
21 reported at the district level and perhaps at
22 the school level, although we might not be
23 able to have the same kind of coverage if we

1 were reporting at those levels than if we
2 could only report at the state level.

3 And part of the reason I guess I
4 have this sort of approach is things that
5 have happened recently, including some of the
6 phrases that we look at in the call -- the
7 Race to the Top information. And even in
8 South Carolina where formerly our -- we had
9 an independent agency that set a 2010 goal.
10 And the former goal was that we would be in
11 the top half of the nation by 2010.

12 The current goal is a totally
13 different kind of vision, as you can see
14 there. And that is that we would graduate
15 students with the knowledge and skills
16 necessary to compete successfully in the
17 global economy, participate in a democratic
18 society, and contribute positively as members
19 of families in communities -- sort of this
20 socialization -- gets to some of the
21 socialization function of education.

22 And even in the Race to the Top
23 call we see internationally benchmarked,

1 college ready, career ready. We see
2 references to eliciting complex responses,
3 higher ordered content and skills.

4 So when we pull together the
5 system it would have a variety of components.

6 And I know I differ from my colleagues on
7 this first one -- or many of my colleagues.
8 I would prefer I think if we have a common
9 set of standards to have a national test, not
10 a myriad of consortium. But remember, of
11 course, that I'm thinking ideally that we
12 would have this reporting at the state
13 level -- so that we would be able to compare
14 states and report state scores in all
15 subjects and so that we could address -- you
16 know, when I think about the future of our
17 society I don't limit it to -- well, if we
18 had only instruction in reading and
19 mathematics I'm not sure we would solve some
20 of the biggest problems that are facing us
21 today or that we would continue to exist in
22 the future.

23 So I would prefer a national

1 consortium. I do realize that we have
2 examples of consortia that have been very
3 effective, like the New England Compact,
4 which works very well with the states in
5 close proximity -- and that may be a key
6 feature there. So why not do states in close
7 proximity like the Secession Compact or
8 Manifest Destiny or even the Trail of Tears.

9 But I don't -- I think that would
10 not serve us well because that -- then we
11 would be fostering -- or possibly fostering
12 different standards now -- across consortia,
13 not across states.

14 And we do have some examples in
15 our English language proficiency examinations
16 where consortia have worked. Those are not
17 identical approaches to the assessment, and
18 maybe that's a good thing. But they're not
19 comparable. And even within the consortia
20 there are different standards so they're not
21 comparable.

22 So if that's what we want is to
23 look across a comparison nationally I think

1 that I would prefer a national test. This
2 would be a summative test. I would prefer,
3 as I said, only that scores be reported at
4 the state level, but certainly not be high
5 stakes for students and teachers. At the
6 very least, you know, we would report at
7 district -- or we could report at district
8 and school but not use this for a teacher
9 evaluation system or focus very heavily on
10 this for students.

11 I do think we should have interim
12 tests which are like mini-summative tests --
13 M-I-N-I -- maybe three to four times a year,
14 that we have a baseline score, and that we
15 report growth scores throughout the year.
16 Now, even though this is something I would
17 like to see this has been problematic for us
18 in South Carolina. We've tried to develop
19 state benchmarks and not been very successful
20 at it.

21 More recently we do have an
22 adoption process, and that has come as a
23 result of offering another opportunity to

1 develop a state interim assessment for
2 districts, and the districts would prefer to
3 select their own. So I do think this has a
4 set of challenges all of its own.

5 These tests could be multiple
6 choice and CR -- and that would go for both
7 types of summative -- the mini and the
8 national -- or end of year test. And, by the
9 way, with these national tests I'm referring
10 mainly to grades three through eight because
11 I do think high school -- some variation on
12 what Carol talked about, throwing in some
13 things that are unique to the United
14 States -- you know, SAT, ACT, IB, and AP
15 emphasis, work keys, some end of course exam,
16 some conglomeration or some relook at high
17 school more in the context of accountability
18 or in addition -- not just the testing, but
19 looking at that as a full accountability
20 instructional system and thinking about all
21 those components would be worthwhile.

22 But back to grades three through
23 eight. I do think that we can use multiple

1 choice and constructive response, but I do
2 think there is kind of an elusive quality or
3 a false promise to the kinds of constructed
4 response and performance tasks that we have
5 on many of our tests. For example, in South
6 Carolina we used to have 25 percent of each
7 of our subjects, including science and social
8 studies, as constructed response tests.

9 And I used to say if you can just
10 give me one good science constructed response
11 item then I would be pleased because it's
12 really difficult to come up with a meaningful
13 exercise that students can do in a few
14 minutes or 15 minutes or even 30 minutes. I
15 mean, how realistic is that, especially when
16 we look toward the future -- toward college
17 and work readiness and life in general. I
18 mean, it may seem that I threw this together
19 in five or 15 minutes, but it really did take
20 me longer to think about that.

21 And so I think on that score we
22 need to move toward extended projects in all
23 subjects, all grades. And these would be not

1 just virtual projects -- these would be
2 projects where students would actually be
3 performing something. It wouldn't have to
4 be -- could be a few hours, it could be a
5 more long-term project.

6 And I think that this is where we
7 need to put a lot of our professional
8 development and our teacher involvement. You
9 know, a lot of times if a student does an
10 extended project and they've put time and
11 effort into it the teacher sort of gives the
12 benefit of the doubt -- doesn't look at that
13 as critically perhaps as he or she should.

14 So I think developing some good
15 extended projects that can be scored by
16 teachers and involving the teachers in that
17 scoring and having training for them at the
18 same time so that they could be trained about
19 the scoring but also hopefully learn about
20 the subject matter as well with, which
21 unfortunately sometimes is necessary.

22 We did something similar to this
23 in South Carolina with one of the West Ed

1 Science Assessments. And those were actual
2 hands-on assessments where students used
3 fairly common items to conduct experiments.
4 And then they were structured in such a
5 way -- so I can see standardizing these -- so
6 that they could be fairly easily scored, but
7 we did find that the teachers were learning
8 the science as well as how to score in this
9 process.

10 So something like that I think
11 would be an important adjunct to what we
12 have, in addition to formative tests, which
13 are the classroom kinds of questioning
14 technique, which, ideally, if we had college
15 programs that were maybe like the Ontario
16 high school programs teachers would learn
17 these techniques in college. But,
18 unfortunately, we do need a lot of intensive
19 training. And we've had some of these
20 intensive projects and they are time
21 consuming and it takes a while to teach -- or
22 to work with teachers in this way.

23 Okay. So as we move on, one of

1 the areas that was mentioned in the call was
2 how would we manage these projects. And
3 there is a lot to think about this in any
4 consortium. The fiscal agency is a mixed
5 blessing. If you are doing it you've got all
6 the work, if somebody is doing it they've got
7 all the control. So who would be the fiscal
8 agent for a consortium? Who would own the
9 materials?

10 There's got to be in my
11 estimation -- having been involved with
12 different smaller collaboratives through
13 enhanced assessment grants, there has to be a
14 strong management plan, as sort of like in a
15 classroom. It's a prerequisite -- classroom
16 management is a prerequisite to learning.

17 And it's the same thing with
18 managing a consortium or a contract. You've
19 got to have somebody to call the meetings,
20 somebody to be responsible, somebody to
21 oversee whether somebody's doing status
22 reports -- that kind of thing. What would
23 happen to people who join late? Do they get

1 the full rights and benefits of everyone
2 else? None of these are barriers, but I
3 think you need to think about all of these
4 things ahead of time.

5 What about state laws and
6 policies? Are you an English-only state? A
7 bilingual state? Do you have separate
8 procurement rules in your state that you have
9 to follow? Do you have requirements for
10 release of items? How about field testing?
11 Are you allowed to field test in your
12 state -- those kinds of things.

13 The decision making -- I'm not
14 going to go through all of these. But just
15 think about all the things that we do as an
16 individual state programs and then magnify
17 that by the time factor that it would take to
18 be in a consortium. Because you can't walk
19 down the hall and consult somebody about now,
20 you know, what was -- what are our
21 measurement guidelines on this? You know,
22 how do -- what is -- what method of
23 psychometric or standards do we use. So you

1 would have to magnify that by some time.

2 Which brings me to something
3 that's close to management, and that is the
4 money involved. And this is named as my dad.

5 His name is Melvin. There's a moose at
6 Disney that's also named Melvin, but it's not
7 my dad.

8 Anyway, this comes -- this is kind
9 of a family joke. Twenty years ago my
10 parents were celebrating their fortieth
11 anniversary and my grandparents were
12 celebrating their sixtieth anniversary so we
13 ended up with a hundred year anniversary. So
14 we have this sort of Melvin math in my
15 family.

16 So if we've got three years to
17 adopt a common core -- and this is sort of
18 reiterating what Lou said -- and then we've
19 got the typical test development process that
20 takes three years, does that equal 24 or even
21 36 months that we have to spend the Race to
22 the Top assessment funds? And if we throw in
23 that magnification factor for consortium --

1 you know, developing a test with a consortium
2 may not take three years -- it may take a
3 little bit longer.

4 If we have \$350 million and we
5 subtract out the 50 percent that's going to
6 the local education agencies, and then we
7 divide that by the 50 states or if there are
8 more jurisdictions clearly that's not enough
9 money -- although it would be welcome money
10 it's not enough money to do all that we want
11 to do in this amount of time.

12 So I guess my last advice -- I'm
13 not sure it's the best -- is to -- don't
14 require this -- this is essentially
15 reiterating what Lou said. Don't require
16 students and teachers and systems to
17 participate in both what's required now and
18 the development of a Race to the Top system.

19 As Carol said, start with the end
20 in mind. And I don't know if we want to go
21 back to the very end -- the democracy, the
22 college readiness, the career readiness, or
23 the common standards -- but I think we need

1 to have some hierarchy of looking at that
2 with the end in mind.

3 For every component and decision
4 try to think about all the possible
5 unintended negative consequences and really
6 brainstorm that. Because I think we are
7 feeling some of the fallout from a very
8 worthwhile system, but the backlash of that.

9 And then consequential validity --
10 this is something that we toss around. But
11 maybe we ought to think from the very
12 beginning -- how could we really assess
13 consequential validity? If we can't sort of
14 design a method for doing that maybe we ought
15 to rethink that.

16 And then something that I thought
17 about a little bit later is -- especially
18 when I was thinking about English language
19 learners and students with disabilities --
20 development is a process, not a thing. So we
21 are learning more about assessment of all
22 types, but assessments especially in these
23 areas. We've seen how much more we've

1 learned, and so we have to somehow not think
2 that there's going to be a final end in mind
3 for that, but that it's going to be a
4 continual process to get better.

5 Okay. And, last, we do have a
6 fine line between rigorous tests that can
7 enhance achievement and defeat and
8 demoralization on the part of our educators
9 and our teachers. We have a fine line
10 between this accountability for all high
11 stakes testing and fostering innovation, but
12 not innovation of the type that we would
13 necessarily want, and unintended
14 consequences.

15 And then the tricky part that I'm
16 not very good about at all is public
17 perception. If you lower -- if you change
18 your standards -- you've lowered your
19 standards -- you know, if you don't report at
20 all of these levels you're keeping something
21 secret. So that's something I think that we
22 need to keep in mind as well.

23 MS. WEISS: Thank you. That was

1 terrific. I think maybe you are our public
2 perception person. You've got lots of good
3 marketing ideas for us on naming consortia
4 and stuff. Questions from folks?

5 MS. POSNY: If I could go back to
6 your grand scheme because I find it very
7 intriguing -- every child with a laptop doing
8 what they should be doing. In the best of
9 all worlds then, if a child was going to
10 receive an assessment at whatever point in
11 time then this assessment system would have
12 to, you know, be able to assess wherever the
13 child might be at that point in time -- if
14 you're following my thinking. You know, is
15 that possible to do within an assessment
16 system -- to have one where you could assess
17 in the -- you know, the fall of the year, you
18 know, before winter break, after winter break
19 and know where they are within the
20 curriculum.

21 MS. SISKIND: Yeah. I think that
22 the summative kind of tests that would be
23 used for comparative purposes across states

1 would be a set point in time. So I don't
2 think that students would be taking that at
3 any -- just whenever they felt ready, at
4 least in the way that I envision it.

5 I do think the interim tests would
6 be set -- also have a set schedule. Okay?
7 Now, it makes it convenient for every student
8 to have a laptop. I do think -- and I'm not
9 exactly sure how to go about this -- but back
10 to what Lou said in terms of some sort of
11 scope and sequence maybe we can -- as we
12 think about our common core standards we can
13 think about the way learning progresses
14 across time.

15 And on those interims, you know,
16 maybe those could be adaptive in the sense
17 that they would go beyond the grade level,
18 whereas I think maybe the summative tests
19 would -- or the national tests would be at
20 the grade level.

21 Also one thing I didn't mention is
22 I do -- I tried to think about where -- at
23 what point -- or what sort of data would be

1 appropriate for teacher performance --
2 assessing teacher performance. And I do
3 think using all of the data that we have at,
4 say, the national summative and the interim
5 data -- whatever we have to make those sort
6 of predictions would be preferable. It would
7 be additional information.

8 I also -- in terms of that
9 component I've -- I do appreciate the
10 approach that South Carolina is taking in
11 that the student performance is only one
12 portion of that.

13 MS. WURTZEL: So I had a related
14 question -- I actually have two questions.
15 One is in your grand scheme how do you see
16 the role of extended projects and whether
17 they are considered in the context of making
18 determinations about school effectiveness?
19 Or how do they fit within this other set of
20 performance data that's being generated by
21 the other assessments?

22 And then you talked about how
23 difficult it's been to generate really high

1 quality constructed response and performance
2 tasks. And I wanted to get your thinking
3 about whether when -- if you had multiple
4 states working together in a consortia you'd
5 be able to put more resources to that problem
6 and generate better tasks.

7 MS. SISKIND: I do. I think
8 that -- well, I think that it -- that having
9 more states work on those more extended
10 projects and even constructed response, if
11 you will -- although I feel that for the
12 typical test -- multiple choice, constructed
13 response -- in general we know how to do
14 that.

15 So I guess I'd like to see the
16 consortium work at the more student level,
17 which would be in your interims, these
18 extended projects, maybe the way that you go
19 about delivering or inculcating teachers into
20 professional development for formative
21 assessments -- certainly want to see some
22 resources put there.

23 I guess I was not thinking -- I

1 was thinking of the extended task as
2 something that would be required of all
3 students, but not necessarily something that
4 would be part of the overall accountability
5 for schools and districts and so forth, other
6 than the fact that they had given -- they had
7 made an opportunity for that to happen. That
8 would be more part of the instructional
9 process.

10 MS. WEISS: So the interims --
11 so -- because I was wondering about the
12 distinction in the system between the interim
13 and the extended project not -- I get the
14 difference in the two types of tests -- but
15 the purposes of them. So for your -- in your
16 thinking one's about accountability and
17 one -- can you just talk a little bit more
18 about those?

19 MS. SISKIND: Well, the interim --
20 I guess I was thinking about I guess as
21 accountability in a sense, but accountability
22 at the student level. So you'd have this --
23 you could see the growth of the student but

1 you wouldn't aggregate those scores for
2 schools or districts or something like that.

3 I think that these extended
4 projects could be a part of some interim
5 system but they wouldn't necessarily have to
6 be a part of some interim system. I think
7 the value of those to the end is the
8 important thing -- I mean, what you want
9 students to know and be able to do in the
10 end.

11 MS. WEISS: And how do you see
12 these fitting together with curriculum? This
13 is sort of on the heels of the question that
14 I think Ann asked Lou earlier. How tightly
15 coupled do curriculum scope and sequence
16 decisions need to be with this assessment
17 process for it to work, and what does that
18 mean for cross state -- or even cross
19 district or cross school consortia?

20 MS. SISKIND: Well, I probably
21 have to answer that similar to the way that
22 Lou did. I mean, we -- again, if we have
23 some sort of common standards that we could

1 view as -- in some sort of learning
2 progression and then our instruction was tied
3 to that -- of course, our assessments tied to
4 that -- we would -- if we give tests at given
5 points in time, whether they be interim or
6 end of year tests, we would have to have some
7 expectation I guess for that -- or at
8 least -- well, I don't know whether we would.

9 At least we need -- the information would
10 tell us where our students were with respect
11 to that.

12 But we probably -- we wouldn't
13 want to hold anybody back and we wouldn't
14 want to force anybody into something too
15 quickly. I think we just need to know
16 accurately where they are in the scheme of
17 things. So I think there would have to be
18 some structure, but I wouldn't see it such a
19 firm structure that we couldn't meet
20 individual needs.

21 MS. WHALEN: On one of your slides
22 you kind of outlined some of the issues to be
23 thinking about in terms of project

1 management. Can you clarify what role you
2 think we should play in helping mitigate some
3 of those areas that could go wrong or whether
4 this is -- with just great advice for your
5 colleagues about as you think about joining
6 together to get answers to these questions?

7 MS. SISKIND: Well, I've been
8 thinking about that over and over and over
9 again. I don't know whether I should propose
10 that you develop this summative national
11 test. I mean, it certainly would -- having
12 some national body with some sort of status
13 do that would certainly help mitigate a lot
14 of this. I think that public perception
15 would be particularly favoring that.

16 So I don't know. I really don't
17 know. I do think that these are deep
18 considerations and that it is money for which
19 you have oversight. So I think you need to
20 take all of these into account when you spend
21 that money.

22 MS. WURTZEL: As Joanne mentioned
23 at the beginning, 50 percent of the funds

1 flow to LEAs. In the system that you've
2 sketched out what do you think about the most
3 powerful uses of LEA funds and what should we
4 be suggesting -- or what would as a state be
5 wanting your LEAs to do?

6 MS. SISKIND: Probably at that
7 formative and extended project level I would
8 see the expenditure of most of those funds.

9 MS. WEISS: Okay. Let's turn to
10 Brian.

11 MR. GONG: Okay. I'm ready to
12 start. I'm going to race through these. And
13 this is aimed at providing advice to the U.S.
14 Department of Education about how they can
15 structure -- I refer to this as an RFP
16 because I think this is -- there may be a
17 different term for it.

18 My main point is that you're
19 shaping what assessment is going to be like,
20 not for one or two years, but for a long
21 time. And so as you shape the RFP which will
22 shape what states and others do you should
23 fund it with a longer term view of having in

1 place dramatically better assessment systems
2 in ten years.

3 And when you have to compromise,
4 which I think you'll have to do, I urge you
5 to choose longer term investments over short
6 term gains. I think you want to say very
7 clearly what you want in the RFP, and to go
8 to as questioned I think you have a role that
9 you can do to help foster good responses to
10 the RFP. I'll talk about each of these.

11 My three main recommendations are
12 I think you should hedge your bets by funding
13 multiple ways to do multi-state common
14 assessments, especially in high school. I
15 think you should invest in a game changer --
16 what I call game changers -- things that
17 will -- that could make assessment
18 dramatically better within a decade but would
19 not be operationally -- could not be
20 implemented operationally in the short time
21 schedule. And I'll talk about why I picked
22 2012 -- and I thought of six of these. And
23 then I'll talk a little bit about fostering

1 this good responses to the RFP.

2 I think that your RFP -- if you
3 have to do it you should fund implementation
4 for the short term, which I'm calling 2012.
5 And that would focus on what we already know
6 how to do -- that is, it's being done some
7 place in large scale assessment -- but it
8 will take time because there will probably be
9 a new set of content standards and it will
10 take multiple states, which has not been done
11 before except for one small consortium.

12 And then for development through
13 2015 I think you should focus on what we do
14 not know how do well at scale but which has
15 potential to lead to dramatically better
16 assessment systems -- and I would separate
17 those two.

18 My point about implementing a new
19 multi-state summative assessment will take
20 years -- and this is if we knew how to do it.

21 And you award the RFP in September of '09
22 the states would get together, they have to
23 do test specification, develop the items and

1 stuff in the next year, they pilot the items
2 in 2010-11, they would have the first
3 operational assessment and reporting in 2011-
4 12. And if they have to use it for
5 accountability that's when accountability
6 would be hit. If you want to do growth the
7 earliest you would have it on this schedule
8 is 2012-13. And if you were using it for
9 high school graduation, which half the states
10 are doing something with, then the class of
11 2015 is the first time you could do that if
12 you implement it on this time schedule.

13 So when people say it's not fast
14 enough, in fact, I don't think you can go
15 faster than this. And all the up-front
16 work -- the aligning the curriculum
17 instruction, accountability, and support --
18 takes longer. But you have to know that in
19 2010 when you're developing your system. So
20 my point is I don't think you can get much
21 innovation in a short -- in something that's
22 going to be implemented for operational use
23 by 2012.

1 So this is just a little side: I
2 think you have to specify what you want, and
3 I think you have to specify how the
4 assessments results will be used. I know
5 that we say it's not for accountability; we
6 don't know accountability.

7 I don't know, as people have said,
8 how to design the assessment without knowing
9 that. If I don't -- if I'm going to have
10 student stakes, then I need to have a certain
11 level of reliability. And if I don't have
12 student stakes, then maybe I don't. And if I
13 don't know that, I have a really hard time
14 developing the assessments. There are lots
15 of things like that.

16 So here are some things I think
17 that you could do very quickly for this 2012,
18 and I have six. The first is if you wanted
19 to hold students, schools, districts, and
20 states accountable to a common performance
21 standard by triggering sanctions, it has some
22 stakes with it. So this is a problem. I
23 said, Well, you know, states are reporting

1 different things from NAEP. What we need to
2 do it get them on the same scale and then
3 we'll do something if they aren't.

4 Then I think that you have to have
5 statistically robust reports of performance
6 on a common metric that is stronger than the
7 current NAEP mapping studies. And I think
8 that means that you have to have a single
9 test. It has to have the same administration
10 procedures, strong equating across the years.

11
12 I don't favor this. I don't know
13 what the theory of action for why we need
14 common state performance standards beyond
15 NAEP. But if someone wants to argue that
16 then I think if you really think it's
17 important I think there's only one way to do
18 it.

19 We know how to do standard space
20 interpretations, but there are several issues
21 that people have brought up. If you wanted
22 to do school -- if you wanted to do state
23 accountability -- we already talked about

1 matrix sampling, what role might that play.
2 In high school we've got lots of different
3 models.

4 So if we wanted to have a common
5 set I think states would have to say are we
6 going to use your model of high school
7 assessment or are we going to use mine? Are
8 we going to use your model of student --
9 assessing students with disability or are we
10 going to use mine?

11 I'll come back to this and form
12 better instruction. I think that there are
13 some things that we could do quickly that
14 would help and form better instruction, but
15 that's primarily I would argue not in the
16 assessment design. I think it's -- as Carol
17 has said there some things in assessment
18 that's primarily outside the supporting
19 systems around it.

20 If we want rapid turnaround
21 because we're trying to promote improvement
22 through the rapid feedback to inform the
23 actions I think we have to say what are the

1 actions and what's the appropriate time
2 frame. There are things that are multi-year
3 or annual and there are things that are
4 shorter term. Most of the things that are
5 shorter term are not -- or traditional
6 summative assessments cannot inform. They
7 don't have the content.

8 So I would -- I'm not a fan of
9 having rapid turnaround of state summative
10 tests because I don't think they inform most
11 short term decisions. And I think that the
12 trade off's high. So if you're going to do
13 this I think we're going to trade speed for
14 quality of the assessment and the cost. It
15 will I think involve reliance on multiple
16 choice and machine scored tests -- and we
17 know how to do this. We will have
18 centralized standardization instead of more
19 complex performances or local scoring.

20 And we -- it doesn't -- the more
21 rapid it gets the more we have to ignore
22 administration variations. So if we're
23 missing students and we want the things the

1 next day we say, Well, you missed it so we're
2 going to have the results without you. So
3 you have to really say that that's important
4 that the timing is worth the non-inclusion.

5 Measuring growth -- I think that
6 there are -- we know a fair amount about
7 measuring growth -- not enough -- and we
8 really don't know how to use growth in
9 accountability. But I think that we could do
10 a fair amount for 2012. We have a lot of
11 issues about how we might do it better, but I
12 think that there are statistical models that
13 people could use. There's not good agreement
14 about which ones to use.

15 Teacher and administrative
16 evaluation -- I don't know very much about
17 this. I think there's probably some that we
18 could do in terms of statistical
19 approaches -- these value added things. I
20 think that there are -- I think that they
21 don't have a proven track record in this type
22 of system.

23 So let me go to these

1 recommendations -- and I'll come back to the
2 first one. So I'm going to talk really
3 quickly about these things I think that would
4 be a great longer term investment. The first
5 is develop technology that provides more
6 evidence of more complex knowledge and
7 skills -- things that we really value.
8 People have talked about this.

9 But I think that we haven't used
10 technology very much. And I think that an
11 investment at this time could really help. I
12 think it's not something that could be
13 implemented immediately, but I think that
14 substantial progress could be made so that it
15 would show up in the next five to ten years.

16 Just as a note about it -- not
17 technology for technology's sake -- a lot of
18 people have talked about this evidenced
19 center design that Bob Mislevy and others are
20 using. I think you have to embed technology
21 within that.

22 So, for example, it's possible to
23 put a bunch of things on a computer screen

1 that students can access. When is it a good
2 thing to do that and when is it not? I think
3 you have to start off with an idea about what
4 you're trying to do and then you design the
5 technology to do it. So if someone says I
6 can let students have an online dictionary
7 that looks at every word, sometimes you want
8 that and sometimes you don't. You really
9 have to know what you want.

10 So start with this evidence
11 centered design, which is what evidence do
12 you want, how are we going to get it, what's
13 enough, how are we going to use it, how do we
14 make use of it, and so on. And then make
15 sure that that's run across -- that is
16 applied to all students. And technology we
17 know has a challenge for doing all students,
18 so just keep that in mind.

19 And the second is sort of the
20 other end. I think if we invested in
21 developing complex performance assessments it
22 would change the nature of assessment in the
23 United States. And you've heard how other

1 countries have this and certainly classrooms
2 have it -- universities have it. We do it
3 when we're not doing accountability and we're
4 not doing summative assessment. There's a
5 lot that can draw on here.

6 But for the complex performance
7 assessments I would do more about specifying
8 the extent of learning and content. You saw
9 that some of what Carol had. They have
10 frameworks for doing -- we don't just have to
11 say that they're performance assessments. I
12 think we know a lot and other people know a
13 lot about what these things actually are good
14 for and how to design them.

15 A second part of the investment is
16 to develop credible administration and
17 scoring. One of the lessons we learned from
18 other performance systems that have come and
19 gone -- and I associate with Kentucky and
20 some others -- is that if there was a
21 challenge to the credibility and operational
22 things. I think that those are solvable
23 problems with a good investment now.

1 Including all students and
2 teachers in performance assessments is real
3 important. I think there's some we know,
4 there's more that we could learn and do that
5 in a short amount of time.

6 I think it's really important that
7 there's a means for certifying the validity
8 and reliability of performance assessments
9 and of combining the resulting data with
10 other evidence. If you talk about with
11 psychometricians they'll say, I don't know
12 how to scale these complex performance
13 things. That's an example of what I mean by
14 combining the information from these.

15 Third is developing local
16 assessment systems, but I frankly don't see
17 how we can develop powerful interim and
18 formative assessments according to my
19 conception without doing curriculum work. I
20 don't think you can say what the students are
21 supposed to -- what they're supposed to know
22 without knowing something about the
23 curriculum. Our content standards are way

1 too thin for that and I don't think the ones
2 on the horizon solve the problem.

3 So I think they have to be linked
4 together. And interim formative assessments
5 are needed to inform learning and teaching
6 directly. I don't think that investment in
7 summative assessments -- how much we do --
8 will ever do that. So this is a third area
9 that -- an investment now doesn't -- I don't
10 think we can do it in the next couple of
11 years. If you said in the RFP give me your
12 best thing I think we would not have very
13 strong models. But we could develop those I
14 think fairly quickly.

15 I mentioned this before, but we
16 need new measurement models and technical
17 criteria for assessments that include complex
18 knowledge and skills. It's great validity.
19 An example is the problem states have had in
20 getting through peer review. Some of it has
21 been a real problem with the quality, but
22 some has been I think that our criteria just
23 are developed for a certain world and may not

1 be applicable to this new type of thing. I
2 wonder whether Ontario could get through a
3 review with their system. That is, we know
4 the current measurement models, assumptions,
5 and limitations, and it will stop innovation
6 if we continue to use those.

7 I think we absolutely have to have
8 better accountability models and support
9 models of how things are going to be used.
10 If we said how -- what is the model for
11 actually helping support change everyone that
12 I know thinks about No Child Left Behind. I
13 think that's a really poor model. But if we
14 say what is a great model and how are we
15 doing it I think we need an investment to
16 help make that happen.

17 And my last one is that I think
18 that an investment for making this a coherent
19 system so that there are specifications from
20 all of those things would be a good
21 investment. So if you ask me, Brian, what's
22 your ideal now -- I actually thought about
23 this -- I don't have one. I think that I

1 don't know enough about this.

2 But I'm pretty sure that I would
3 say something, as Teri was trying to say,
4 that the state level assessment and the cross
5 state assessment has a very different design
6 than the state level assessment. I think
7 doing -- assessing every study so that we can
8 report state performance is a really
9 inefficient design. I think we've learned
10 that before and I hope we wouldn't have to
11 repeat that type of learning -- so having
12 something at these different levels and what
13 type of information and so on.

14 Okay. So I have these six areas.
15 So let me go into this one about hedging
16 bets. So if you're going to do things that
17 happen in 2012 and it's around common
18 assessment, college/career readiness,
19 especially high school, I would look at the
20 good current models. I think there are many
21 of them -- we've heard people talk about
22 them. They're tightly interwoven with state
23 policies. As I mentioned over the half the

1 states have high school exit requirements
2 tied to testing, and unwinding that is really
3 hard.

4 So what I would do is just look
5 for an incremental improvement on that, and I
6 would encourage you to say, let's look at a
7 lot of these and not try to pick just one.

8 And so if you fund multiple
9 versions of this -- and I would even
10 encourage to you think about multiple common
11 content standards to help find out what the
12 costs and benefits of these are. I think
13 it's an unproven model. I don't think that
14 we have one set of content standards that's
15 clearly superior. The one that everyone's
16 talking about, the NGA CCSSO, is not done
17 yet. I think that adopting it without
18 knowing what it will really turn out to be is
19 likely to be a bad thing.

20 So I know we're in a tough
21 position but I would -- I think that the
22 possible downside is really high on that.

23 So you may end up with a portfolio

1 of awards that have strong models that
2 represent incremental advances that can be
3 implemented strongly by 2012 that help get to
4 the longer term goal. But -- and then I
5 would encourage you to think about multiple
6 awards that really develop these advances and
7 try to make them so that they are strong
8 enough for people to say, I really want to do
9 that no matter whether I get this money or
10 not.

11 The other thing is in terms of
12 strategy is you might want to think --
13 because I was thinking about if this is an
14 incentive there's incentives and there's
15 rewards. You know, it was talked about these
16 are rewards for strong performance. If I'm
17 going to do that then I'm going to reward
18 states that are already going to do
19 something. If it's incentive I may fund
20 things that wouldn't happen otherwise.

21 So if there are strong models that
22 are going to happen anyway, maybe you don't
23 fund them. It's really unpopular, but

1 it's -- think about what you're doing with --
2 whether this is incentive or not.

3 Okay. So just the last part about
4 fostering responses: In my experience with
5 RFPs and complex projects, it's not enough to
6 tell people they're deliverable. If you want
7 people to work together, you have to promote
8 the leadership to make that happen. I think
9 there are states, NGOs, such as Achieve and
10 others, test vendors that this has to work
11 in -- I think having some models and actually
12 some support to encourage that working
13 together would go a long way to help make
14 this happen.

15 If you just publish the things,
16 there will be some coalitions that form --
17 that are already forming. I'm not sure that
18 they will end up with a full set of strong
19 models that you would like.

20 So I think you have to provide
21 clear RFP specs for that. If you want states
22 to have vendor partners in their RFP
23 responses, you better tell them that now, and

1 you better think about what those coalitions
2 are. This is a very important point. If you
3 say the RFP means states have to do it -- I
4 don't want states issuing an RFP after
5 this -- better tell them that now. Otherwise
6 they're going to be saying -- the work will
7 come in September and they will issue an RFP.

8 If you want them to have a partner
9 already, better tell them. And that's a
10 really -- and think about what those likely
11 coalitions will be. Who will partner with
12 whom if that's the requirement that you have.

13 Think about what states who don't get the
14 money will do and think what will happen
15 after it's adopted and what will shape things
16 in the future.

17 And here's just a little thought
18 piece. What happens -- what would you think
19 if this were the result? 2012 there were
20 five widely used assessments; they were all
21 aligned to the common content standards.
22 Four were commercially available from current
23 test publishers; that is, you could buy

1 something like you could now with the Achieve
2 Pearson Algebra II -- and one was available
3 by joining a consortium like WIDA.

4 And the states were purchasing one
5 from one and another from another and mixing
6 them. What would -- so is that good or bad?

7 What if there were only one assessment?
8 What if there were 46? Okay. So I think you
9 can shape that by how you shape the RFP.

10 MS. WEISS: Thanks. I feel like
11 starting by asking you to answer that last
12 question. Yeah, what do you think about
13 these different scenarios? Which ones are
14 good and which ones --

15 MR. GONG: I think that there are
16 going to be commercially available
17 assessments no matter what you do. I think
18 you have to think about that and think about
19 whether that's all you want or whether you
20 want something in addition to it. I don't
21 think you can stop vendors from doing it. I
22 think that they are likely to be the leaders
23 for many of the consortia anyway.

1 So I think that that -- if that's all it were I
2 think that that would be a problem for me.
3 And partly because in the last 20 years
4 states have done a terrific job of innovating
5 and coming up with things. And they work
6 with their partners.

7 But if you think about the
8 diversity of high school assessment, that's
9 happened because states have figured out how
10 to do these different things. And when we
11 start to close down, I worry about not that
12 these won't be good, but I worry about
13 where -- what will happen in the next
14 generation. And so that's partly what I mean
15 of think about the structure that's being set
16 up that will shape how will the next
17 generation of assessments are actually going
18 to -- how they're -- those are going to
19 arise. So I worry about having only
20 commercially available ones that people buy
21 into.

22 MS. WEISS: And so what's the
23 ideal scenario for you look like five, ten

1 years from now?

2 MR. GONG: I think in five years
3 from now -- I think that we can keep going
4 and have tests that serve the current
5 purposes. I think that there can be
6 incremental improvements on that. I think in
7 five years from now I think what we're doing
8 is starting to get into tests that have these
9 other characteristics of merging performance
10 and --

11 MS. WEISS: Right.

12 MR. GONG: -- technology from the
13 design.

14 MS. WEISS: But I mean from a
15 market point of view. How does states get
16 those and do you have to be in a consortium
17 to use them or can you --

18 MR. GONG: I think that --

19 MS. WEISS: -- get them if you're
20 not in the -- what's the marketplace look
21 like --

22 MR. GONG: Yeah.

23 MS. WEISS: -- assuming that

1 publishers are going to do whatever they're
2 going to do regardless --

3 MR. GONG: Yeah.

4 MS. WEISS: -- of where these funds
5 come down. If there's a common set of
6 standards that's going to create a market for
7 publishers.

8 MR. GONG: I think that states
9 that have a clear vision will have the --
10 will get the money to make that happen
11 anyway. So I think about Colorado with their
12 growth model. They're using some federal
13 funds, but they are committed to doing that.

14 And Massachusetts is doing what they're
15 doing. I think that will be a market where
16 states that want to make things happen will
17 help make it happen.

18 I think that part of what's
19 happening now with the whole breaking open of
20 common things is that there will be a market
21 where states are going to start to share much
22 more. I think that will be a lasting -- like
23 I said, I think that will be healthy. It

1 will create more options for people to get
2 things that they couldn't afford otherwise.

3 MS. WEISS: Questions?

4 MS. POSNY: Yeah. It's
5 interesting for me to -- you know, for you to
6 note that you're not necessarily recommending
7 that one common core sort of standards. I
8 mean, I think the states have finally come
9 together to say, okay, let's take this piece
10 off the table. Let's get the summative --
11 let's just get that one.

12 With the idea then that then the
13 innovativeness and all the rest would be all
14 the other things we're talking about -- the
15 performance assessment, the project based
16 learning. I mean, why are you thinking we
17 still need multiple -- you know, it hasn't
18 worked to have 50 different sets of state
19 standards nor 50 different assessments. So
20 I'm just interested in knowing why you think
21 we still need multiple variations on a theme?

22 MR. GONG: Actually, I would say
23 that it has worked, you know, pretty well in

1 certain ways. But there have been maybe
2 three areas that we could improve on. We
3 know that the comparisons across states is
4 not a very happy one. However, I think that
5 you can use NAEP to compare state
6 achievement.

7 You don't have to have -- and
8 people -- all these mapping studies show that
9 you don't have to have common content
10 standards nor a common assessment to put
11 states on a common metric. We don't need
12 that in order to do the common comparison.

13 Now, if we wanted to have a tight
14 common comparison, we would need to have a
15 common assessment. So if that's really
16 important to us, then I'd say that we ought
17 to go ahead and do that.

18 The second reason for having
19 common content standards that people have
20 said is that the current quality of states'
21 content standards is not very good. And we
22 can -- and they won't change it on their own
23 so let's create some and send for them

1 adopting.

2 So I think that will have -- if we
3 had something that was dramatically superior
4 I would agree. I don't think we have that
5 yet. But let's wait and see the evidence on
6 that.

7 The third is efficiency. And
8 there are a lot of good arguments. You
9 notice I left efficiency out here. Actually
10 freeing up money and time may actually change
11 the nature of what states could do in a
12 really healthy way. Lou was saying, I
13 thought the reason for having common
14 assessments was so that we wouldn't have to
15 have 50 peer reviews. I think states would
16 welcome that.

17 So, you know, there -- what I'm
18 trying to put here is that if the quality is
19 high then -- which I think we're assuming --
20 then there may be many advantages. If the
21 quality is not high -- and that's what I was
22 trying to point out by this 2012 -- I think
23 that it's very likely that the quality is not

1 going to be that much better in terms of
2 implementation than -- I think it -- I'm
3 afraid that by rushing into it, it will close
4 down the options and benefits that we may
5 have.

6 MS. WURTZEL: So, Brian, as I
7 understand it you're describing two work
8 streams. One is incremental improvement of
9 current state assessments, potentially by
10 consortia of some size, coupled with the game
11 changers.

12 So my question to is should --
13 from the federal level should we be funding
14 both work streams? What's the relative
15 importance of them? And how eventually do
16 they come together into one coherent set of
17 work?

18 MR. GONG: I think that's exactly
19 the right question. Let me go back to this
20 one about incentives. Some things will
21 happen -- would happen anyway. And we
22 noticed in Kentucky people said, Don't you
23 care about physical fitness? You haven't put

1 basketball in your school accountability
2 system. I said, you know, we don't have to
3 put that in the school accountability system
4 for people to pay attention to basketball in
5 Kentucky.

6 So the important point is are you
7 doing an ideal system or are you doing
8 incentives to help create things? It's a
9 really important question. If you're doing
10 an ideal system, then you better fund
11 everything that you think is important. If
12 you're -- and I don't think you have enough
13 money for that. If you're doing incentives,
14 you're trying to help make things happen that
15 wouldn't happen or wouldn't happen as
16 quickly.

17 So I think that there are ways for
18 states to get states to work together that
19 may not take this type of money. But I think
20 that the game changers will not happen
21 without your investment.

22 VOICE: How do they come together?

23 MR. GONG: I think there are two

1 ways to have them come together. I'm a fan
2 for having validity in things. States will
3 recognize quality. If we said here are some
4 ways that are technically feasible --
5 operationally feasible for performance
6 assessments to be brought in. I think many
7 states would say, if you show it to me -- how
8 I can do it then I'll do it.

9 So these -- and part of the
10 challenge for these game changers is how to
11 make them operationally feasible. We already
12 know how to do performance assessments on
13 small scale. How do you make it large scale
14 and credible? That's what you're investing
15 in. You're not investing just in the
16 development. So the first one if you have
17 the quality and the operational feasibility I
18 think states will voluntarily come in.

19 But then I think that there are
20 ways to -- if we're still concerned about
21 commonality that there are ways to create
22 incentives. There -- and one model is --
23 your Race to the Top is all volunteer. No

1 Child Left Behind was all mandatory. I think
2 thinking through that mix would actually give
3 you some other tools.

4 MS. WHALEN: One of the
5 questions -- or one of the concerns you
6 raised was about our peer review process and
7 just how bad as stifling quality assessments.
8 And it maybe I'm overstepping what you said,
9 but what is your thinking about how that
10 process should look like?

11 MR. GONG: Well, I have all these
12 states here -- they could tell you about it.

13 The peer review process purpose is to ensure
14 fundamental quality. I think it's a -- it's
15 sort of like a minimal thing. You're
16 certified to practice. We're not saying
17 you're a great doctor; we're just saying that
18 you're not going to endanger public welfare.

19 So I think having that mind set
20 would be helpful. Now, peer review is
21 really -- I think in the long term has helped
22 many states for what they've done. There are
23 two specific things. One is, for the things

1 that we don't know about I'm not sure that
2 inventing the criteria after people are
3 working is a great process. Okay? So that's
4 just part of knowing what we want up front.
5 And that's being balanced, of course, with
6 some improvement.

7 I think the other is a similar
8 one. In the overall scheme of things the
9 idea about withholding full approval means we
10 have to have some leverage on you or you
11 wouldn't do it. That is, if this is an
12 incentive it may not be true. The states may
13 be willing to do those things without having
14 that -- without having full approval or going
15 through that complex process. They may be
16 willing to do it anyway.

17 So think about what the real
18 purpose is. This is this theory of action.
19 You know, if we're trying to -- if it's a
20 certification step that's one thing. If it's
21 an incentive step or a motivation thing
22 that's another. And I think being really
23 clear about what that is would help.

1 And so the specific thing -- if
2 there's a minor thing that states are working
3 on -- and there are states that are working
4 on this 1 percent assessment. They are going
5 to get it right -- those are really tough
6 issues. I don't know any state who says if
7 you gave me approval to move ahead I would
8 quick work on it. I think they're all
9 dedicated to doing it. But it's sort of a
10 big process for a -- at this point in time
11 anyway -- a relatively -- something that is
12 relatively small but states are already going
13 to do I think.

14 MS. WEISS: Great. Let's take a
15 lunch break. We'll come and hear from Tom
16 and Paul after lunch. I think we will
17 reconvene in this room at 1:15 if we can.
18 And, remember, if you've got any note cards
19 that you wrote us comments on drop them at
20 the front desk on your way out so that -- at
21 the registration desk on the way out so that
22 we can grab those cards before the roundtable
23 starts. And thank you so much. We'll see

1 you back here around 1:15.

2 (Whereupon, the meeting was
3 recessed, to resume at 1:15 p.m.)

4

5

6

7

8

9

10

11

12

13

A F T E R N O O N S E S S I O N

1:20 p.m.

MS. WEISS: So thanks. I think we are going to start with -- or resume with Tom Fisher. So, Tom, if you are ready we'll turn it over to you.

MR. FISHER: Thank you. I'm pleased to be with you today to discuss how states can respond to the requirements of the Race to the Top initiatives authorized under the American Recovery and Reinvestment Act of 2009. This program seeks to encourage states to improve their statewide student assessment programs with the aim of better preparing students for success in post-secondary education and in the 21st century workforce. This is a significant undertaking.

There is no single way to accomplish this objective. Each proposed new solution of approach will have both intended and unintended consequences that will impact your state's educational

1 programs. You will need to proceed with due
2 diligence.

3 I have been in the large-scale
4 assessment business for 40 years at both the
5 local and state levels and have learned a
6 few lessons along the way. Based on these
7 experiences I want to offer a few ideas that
8 may help you in your planning.

9 Statewide assessment programs are
10 designed to fulfill specific policy
11 directives and purposes. One does not
12 develop a test and then seek purposes for
13 it. You first determine what kinds of
14 decisions need to be made about students,
15 schools, and systems and then design a
16 testing system that will produce scores to
17 facilitate those decisions.

18 You should not make statements
19 that assign additional purposes unless you
20 have conducted research needed to validate
21 such new purposes. For example, building a
22 high school test that allows valid
23 assertions about probable success in the

1 first semester of college does not imply
2 that the same test will predict who will
3 actually graduate from college or who will
4 become a good employee.

5 Developing a statewide assessment
6 program where one previously did not exist
7 is actually fairly simple compared to
8 redirecting an existing program. The latter
9 requires much more effort and planning.
10 Events must be carefully sequenced so there
11 will be no break in services or in the data
12 produced by the tests.

13 This is particularly important if
14 you presently have a high-stakes testing
15 program and want it to be replaced with a
16 new, more challenging test which also has
17 high stakes for the students. Any student
18 currently working toward an existing high-
19 stakes diploma must be permitted to continue
20 until that diploma has been awarded. The
21 state, therefore, would have to operate both
22 types of assessment programs for four or
23 five years. There obviously will be cost

1 implications.

2 If the proposed common standards
3 assessment program will have high stakes at
4 the student level it must adhere to the
5 legal principles established through prior
6 litigation such as in Florida's Debra P.
7 versus Turlington and the G.I. Forum case in
8 Texas. Students have a property interest in
9 receiving a high school diploma and this
10 right cannot be ignored.

11 At the very least, this means
12 that students must understand when they
13 enter ninth grade what the academic
14 expectations will be, and the state must be
15 able to demonstrate that all students were
16 provided the opportunity to learn the
17 required content.

18 Additionally, in some states test
19 security is an issue since a student is
20 provided the opportunity to see the test he
21 or she failed. This means that there must
22 be alternate forms available for each re-
23 test administration.

1 If the assessment program is tied
2 to high stakes only at the school or
3 district level there are no personal legal
4 rights to honor, but there certainly are
5 practical and political considerations.
6 Educators surely must believe the standards
7 are fair and equitable for all students.
8 Other interested parties will want
9 assurances that the standards are
10 sufficiently challenging to accomplish the
11 goals of Race to the Top.

12 This is a difficult balance. No
13 one will be happy if, under a new testing
14 program, all schools earn failing grades.
15 Or if all schools achieve the highest
16 possible evaluations it would not lend
17 credibility to the assertion that more
18 challenging expectations really are being
19 implemented.

20 It is appropriate for us to have
21 high expectations for students, but some
22 students and their parents may not believe
23 that post-high school education is

1 necessary. The local culture may very well
2 believe that the possession of a college
3 education means the young man or woman will
4 move away from home, never to return. This
5 breaks up the continuity of the nuclear,
6 multi-generation family structure.

7 I believe this dilemma can be
8 solved by authorizing differentiated high
9 school diplomas. These diplomas would
10 include a minimum skills level of
11 performance, a higher career or technical
12 diploma, and a college or university
13 proficiency diploma. Students would move
14 towards whichever one they desire and should
15 be allowed to change their selection at any
16 time if they wish.

17 There are absolutely no
18 proficiency standards or cut-scores that are
19 absolute and can be applied to statewide
20 assessment tests. We have developed very
21 good approaches to setting criteria for cut-
22 scores, and they are available to be applied
23 to the next generation of tests.

1 There always is a question as to
2 whether a given test is too hard, too easy,
3 or just right. Some states have had
4 difficulty in increasing the rigor of their
5 current tests. In fact, within the last few
6 weeks there have been new stories from three
7 different states describing how attempts to
8 increase academic standards have not been
9 successful.

10 A possible solution to this
11 situation is to adopt cut-scores along with
12 a requirement that they be reviewed every
13 two years with the intention of increasing
14 the rigor so as to maintain forward
15 momentum.

16 Do not assume that classroom
17 teachers understand principles of
18 psychometrics or good test design. Teachers
19 are not required to take courses in testing
20 and measurement during their undergraduate
21 teacher preparation training. There is no
22 reason to assume that all teachers know how
23 to write good test questions for use in

1 their own classrooms, much less write items
2 for a statewide assessment.

3 On the other hand, they can be
4 taught how to do these tasks, and we have
5 many examples of talented teachers who can
6 produce good test items or can serve as
7 graders of students' responses to
8 constructed response items. If it is
9 desirable to include more sophisticated
10 stimulus material and assessment strategies
11 in the next generation of assessment and
12 classroom instruction it will require
13 significant pre-service and in-service
14 training.

15 I now turn my attention to
16 suggestions for your consideration as you
17 think about a next generation assessment
18 program. For purposes of these remarks I
19 will assume the adoption of a new assessment
20 program for grades three through eight in
21 reading, language arts, and mathematics in
22 accordance with Race to the Top.

23 I will also assume that states

1 will provide tests at the secondary level
2 for the purpose of measuring progress toward
3 college-ready graduation standards. These
4 tests may or may not have high stakes
5 associated with them. Remember, as a side
6 note, that at this time about half of the
7 states use high stakes high school
8 graduation tests.

9 At the high school level each
10 state must make a choice between using
11 traditional tests of reading, writing, and
12 mathematics or switching to specific end of
13 course tests. The general tests are much
14 easier to design and implement.

15 By contrast, using end of course
16 tests, usually given via computers, creates
17 difficulties with scheduling of test
18 administration late in the term, field
19 testing new items, instantaneous grading,
20 selection of equated passing scores, and re-
21 test opportunities. The determination of
22 passing scores for each new version of an
23 end of course test is particularly

1 problematic.

2 There is insufficient linkage in
3 most states between high school expectations
4 and college entrance expectations. Student
5 course grades in high school may not have
6 much validity and can be artificially
7 inflated. As a parenthetical note, I note
8 that my local high school had eleven
9 valedictorians in 200, and a school in
10 California boasted of 59 valedictorians. As
11 USA Today commented, quote, Being first in
12 your class ain't what it used to be.

13 Earning a passing grade may not
14 mean the student actually has sufficiently
15 mastered the content. Moreover, across the
16 high schools within a single district the
17 teachers of, for example, Algebra II
18 probably have different grading standards
19 and content expectations. This problem is
20 compounded by variations in course content
21 and grading standards across all the high
22 schools in the state.

23 I would recommend that each

1 district superintendent be required to
2 certify that all the mid-term and final
3 examinations in specific English, language
4 arts, and math courses be coordinated so
5 that one high school does not have easier
6 standards than another. Implementation of
7 state end of course tests will guarantee
8 consistent educational quality only if
9 students are actually required to pass with
10 no alternative pathways to earning the
11 course credit.

12 I also recommend that if your
13 state really wants to prepare students for
14 success in the post-secondary world the
15 development of such tests and setting of the
16 passing scores should involve college and
17 university teachers, analysis of the current
18 content expectations for college freshman
19 English and math courses, and long-term
20 tracking of student success.

21 The Race to the Top literature
22 envisions high-quality assessments and an
23 overall system that includes formative

1 assessments, interim assessments, and
2 coordinated instructional improvement
3 systems, all structured to permit fast
4 turnaround of test results. Student growth
5 over time will be monitored and linked to
6 the performance of individual teachers. I
7 summarized this in two sentences, but the
8 implications are enormous.

9 Let us first imagine that there
10 will be a summative instrument administered
11 via computers, scored promptly, and results
12 available in a matter of hours. A system of
13 this nature will require a large and varied
14 item bank, a sophisticated computer system
15 for administration, rules for retesting
16 students who fail, and up-to-date equipment
17 in all schools. It is doubtful that such a
18 system could include any items that are
19 scored by human readers.

20 Now, consider the matter of
21 interim tests. In one scenario there may be
22 one test administered at mid-year, subject
23 to the constraints just mentioned. Or there

1 might be two interim tests, one at the
2 middle of each semester. Again, if
3 immediate turnaround of results is demanded
4 constructed response items probably could
5 not be included. If we assume that the
6 purpose of the test is to determine if the
7 student is on track for the end of the year
8 test it will be necessary to validate the
9 degree to which a given score can predict a
10 level of performance on the final assessment
11 at the end of the year.

12 The Department of Education
13 envisions the development and implementation
14 of a series of formative assessment
15 activities to be administered in conjunction
16 with the instructional program throughout
17 the school year. These activities could be
18 computer administered, but they also could
19 be paper/pencil tasks or small group
20 projects graded by classroom teachers.

21 The immediate question to be
22 asked is how will teachers be involved in
23 the creation, administration, and scoring of

1 these assessments? Will this require
2 additional work time for which they must be
3 compensated? What will their training needs
4 be? And what will be the relationship of
5 these activities to whatever content is
6 included in the ordinary offerings of
7 textbook publishers?

8 The continued administration of
9 summative, interim, and formative
10 assessments will require an enormous number
11 of test items of all types, especially if
12 some items will be annually released to the
13 public. It is not clear that these items
14 can all be gathered from existing sources,
15 so new and continued item development will
16 be required.

17 To support such efforts I would
18 create within each participating state
19 assessment development and training centers
20 associated either with universities or large
21 school districts. These centers can be used
22 to review, edit, and validate new test
23 items, write new instructional materials,

1 and train teachers, principals, and
2 instructional leaders as needed.

3 There is no question that modern
4 computer and web technology enables the
5 administration of instructional programs as
6 well as assessment exercises. But it is not
7 clear that all instructional objectives can
8 be measured via computers, such as working
9 in groups.

10 Further, there always is a
11 question about the adequacy of bandwidth and
12 the availability of a sufficient number of
13 computers within individual schools so that
14 all students can be tested within authorized
15 windows. If computers are not available or
16 if some students cannot take the tests in
17 this way, alternate testing procedures must
18 be created.

19 Technology continues to improve,
20 even as we meet today. I noticed last week
21 that Microsoft will be offering software
22 that will permit several students to link
23 their keyboards and monitors to a single

1 classroom computer in a type of mini-
2 network. This should expand the number of
3 work stations without having to purchase
4 more computers.

5 The assessment system being
6 envisioned by Race to the Top requires
7 thorough and prompt reporting of results.
8 Depending on the way in which the computer
9 tests are scored the results could be
10 quickly available. Results aggregated at
11 the school district and school levels will
12 require additional time. The district and
13 state results should be readily available to
14 the public for viewing or the downloading of
15 research data files. Currently not all
16 states have web sites that offer such
17 services.

18 I have said nothing about the
19 continued use of nationally normed
20 achievement tests. Many parents will want
21 to know how well their child performed
22 compared to national norms. If Race to the
23 Top leads to a substantial number of states

1 using the same summative testing apparatus
2 de facto norms will be available and can be
3 used to describe a student's performance
4 relative to other students, although this
5 distribution may not be, quote, national,
6 unquote.

7 Otherwise, to provide national
8 normed data it will be necessary to either
9 separately administer such a test or to link
10 the Race to the Top tests to the scales of
11 an available norm-referenced test. Again,
12 there are cost implications.

13 I now turn to a few managerial
14 matters. States should track all high
15 school students after they graduate and move
16 toward employment, college enrollment, and
17 graduation with a post-secondary degree. If
18 you want to see an example of how this can
19 be done I recommend that you look at the
20 Florida FETPIP -- that's spelled F-E-T-P-I-
21 P -- it's obviously an acronym -- FETPIP
22 operation. You can find it on the web.

23 Initial development of the new

1 challenging assessment program will be
2 supported by Race to the Top funding. It is
3 not clear what the long-term fiscal impact
4 will be on states that implement the new
5 academic standards and assessments.

6 I recommend that you include
7 budget analyses and expenditure forecasts so
8 you will be adequately prepared. New
9 assessments that feature constructed
10 response items will be costly and, indeed,
11 some states presently are minimizing or
12 eliminating such questions because of the
13 costs and time required for grading.

14 The Race to the Top program is
15 encouraging states to develop consortia to
16 develop the needed assessment instruments
17 and procedures. This can be a means of
18 spreading the costs and gaining economies of
19 scale. It will be necessary for you to
20 determine whether or not there are any laws
21 preventing participation in such
22 consortiums, negotiating ownership of the
23 resulting products, building a management

1 team, deciding how to parcel out the work
2 activities, and agreeing on the many details
3 of the work.

4 Maintenance of a tight work
5 schedule will be absolutely required.
6 Consideration must be given as to whether
7 the consortium will be used just for initial
8 development and administration or if it will
9 be an on-going commitment for test renewal,
10 administration, scoring, and reporting.
11 Will your state permit such an obligation of
12 funds into future years?

13 Race to the Top allows states to
14 adopt common academic standards, but each
15 state may have up to 15 percent additional
16 standards by content area that may be
17 unique. If each state decides to measure
18 additional content there could be 50
19 different analyses of data, item sets,
20 equating processes, and so forth.

21 Alternatively, each state could
22 assess its additional content by
23 administering separate tests with results

1 not reported on the same scale as the common
2 national test. National comparison data
3 obviously would not be available for those
4 items.

5 Finally, we must understand that
6 each state already has existing contracts
7 for assessment and school accountability
8 services. Typically, these contracts have
9 initial terms that range from two to five
10 years with several optional renewals. The
11 contracts were the result of competitive
12 bids in order to get the best possible
13 pricing.

14 To change direction for the
15 existing statewide assessments means the
16 states will have to negotiate contract
17 amendments without the benefit of
18 competition or must cancel the existing
19 contract, issue a new RFP, and select a
20 contractor to operate the newly developed
21 assessment program.

22 As I began my remarks today I
23 said that the Race to the Top initiative

1 presents a complex set of challenges. As
2 you move forward I urge that you review
3 every aspect of the state's approach to
4 curriculum requirements, instructional
5 programs, promotion and graduation
6 requirements, the design of the present
7 testing programs, and the school
8 accountability programs so you will
9 understand the adjustments and changes that
10 must be made to move forward in meeting the
11 Race to the Top objectives. Thank you for
12 your attention.

13 MS. WEISS: Thanks, Tom.

14 Questions?

15 MS. WURTZEL: Tom, do you -- very
16 quickly -- because you spoke very quickly
17 for this entire presentation to put a lot
18 in -- went over the issue of quick
19 turnaround and suggested that there are lot
20 of costs associated with requiring quick
21 turnaround of assessment data. So could you
22 elaborate on that a little bit more and
23 speak about whether there are some things

1 that should be quick and some things that
2 could be slow? Or does everything have to
3 be on the same time line?

4 MR. FISHER: Everything does not
5 have to be on the same time line. It
6 depends upon how you set up the program.
7 For example, it's not unusual for a state to
8 say we value having students write. So
9 we'll have them write in February, it will
10 be processed in March and April, and those
11 results merge together with the results of
12 the multiple choice test later on.

13 Now, if you wanted to have
14 extensive math problems that also required
15 those kinds of extensive solutions you could
16 put it on the time line. But then that
17 immediately raises the question in terms of
18 mathematics of the instructional timing and
19 whether or not the student has actually been
20 taught that which he's being measured at
21 mid-year. With writing an essay it's not
22 quite the same thing.

23 In end of course tests it's a

1 whole different kind of problem because you
2 have to worry there about how am I going to
3 set the passing scores if I don't have real
4 good field test data upon which to set cut-
5 scores. It's a whole different set of
6 complexities.

7 MS. POSNY: I've got just a
8 couple of quick questions. One is you
9 talked about reviewing or possibly changing
10 the cut-scores every two years, you know, as
11 a recommendation in terms of how well
12 they're doing. But what about trend?

13 MR. FISHER: The problem we have
14 with setting cut-scores on any tests if
15 people think that they're absolutes. There
16 is no end to how far we can push the
17 envelope. The educational system needs
18 dramatic improvement over a long period of
19 time. And the only way you can get
20 political buy into these things is when the
21 people who have to vote on it are willing to
22 say we're going to keep pushing the
23 envelope. And most people don't want to do

1 that.

2 I remember one time in my past
3 where I thought I'd get around that by
4 having -- by proposing a set of graduated
5 increments in the passing standards in the
6 high school graduation test at two years,
7 four years, six years. We never got to the
8 six years. Everything was changed before
9 they got to the four years.

10 So I think you have to make an
11 overt commitment to say we are going to
12 review whatever the cut-scores are every two
13 years. And then you can either vote to
14 continue the existing level or you can vote
15 to move ahead. But somebody's got to force
16 the issue.

17 MS. POSNY: One other one that I
18 wanted to ask, because I just want to
19 clarify because I wasn't quite sure if I was
20 following you. You talked about the fact
21 that different school in district follow
22 different standards? And I'm not quite
23 following you because the way the assessment

1 system is set up right now you -- it would
2 be folly for any district -- I mean, any
3 school within any district in any state not
4 to follow the state standards.

5 MR. FISHER: I was referring
6 to -- nothing related to the test. I was
7 referring to the fact -- you take any large
8 urban area -- there might be ten high
9 schools -- 15 high schools -- and they all
10 have Algebra I. And they do not share
11 common expectations.

12 So what I was referring to was as
13 a step toward moving people in the direction
14 of thinking of shared commonality of
15 standards make the superintendents as a
16 local matter make their teachers sit down
17 and agree upon what Algebra I should be.

18 MS. WEISS: Other questions?
19 Okay. Let's turn it over to Paul.

20 MR. NICHOLS: Thank you.

21 MS. WEISS: Hang on. Your slides
22 are coming.

23 MR. NICHOLS: Okay.

1 MS. WEISS: Here they are.

2 MR. NICHOLS: Thanks. I was
3 describing this talk to my -- at the dinner
4 table the other night to my wife and my 15-
5 year-old. And exactly -- I hope that you
6 guys don't have the same response that my
7 15-year-old had, which was to roll the eyes,
8 loud sighs, shake of the head. So I'm
9 hoping for a better response here.

10 So I'm going to cover really
11 three different areas in my talk in these 20
12 minutes. First I'm going to invite you to
13 think differently about assessment. I'm
14 going to invite you to think differently
15 about the distinction between formative and
16 summative assessment, about the attitude
17 that educational measurement is quantifying
18 something. And I'm going to invite you to
19 think differently about the definition of an
20 assessment as a set of items in a test form.

21 I'm also going to present a
22 vision of an assessment system as not an
23 assessment but a system for learning -- that

1 is, a system that coordinates assessment and
2 instruction and professional development all
3 at the same time.

4 I'm also then going to finish up
5 by talking about some of the technical
6 qualities of this vision -- reliability,
7 validity, vertical alignment -- some of
8 those sorts of technical issues that as
9 someone who hangs out with psychometricians
10 I'm concerned with a lot.

11 So let's talk first about the
12 distinction between formative and summative
13 assessment. You know, we just -- my
14 colleagues and I just published a paper in
15 educational measurement issues and practice
16 in which we said there's no such thing as a
17 formative assessment. It's not the
18 assessment that's formative -- it's the use
19 of the information makes it formative.

20 So you can give a test that's
21 supposed to be formative -- that's
22 advertised this is a formative assessment --
23 doesn't help your teachers at all -- it

1 doesn't help students learn at all in your
2 particular
3 context. So it's information that's used
4 for formative purposes. It's not the
5 assessment that's formative.

6 The assessment by itself I like
7 to say is like one hand clapping. You need
8 that other hand if you really want to make
9 any noise. That other hand is the
10 instructional part. You have to have
11 instruction that is aligned with the
12 assessment information to actually impact
13 and increase student achievement.

14 So assessment information must
15 sit as a component of a system of assessment
16 information and instructional actions that
17 are coordinated. So if the teachers don't
18 have any way to use the information -- or
19 consume the information that's coming from
20 an assessment and put it into action it is
21 like one hand clapping. It is like that
22 tree falling in the forest where no one
23 hears it.

1 The next -- the last thing I want
2 to leave you with is that information from
3 the same assessment may be used for
4 formative or summative purposes. You don't
5 have a formative assessment or a summative
6 assessment.

7 For example, a lot of you are
8 familiar with mastery learning. In mastery
9 learning your students will take a test.
10 The initial assessment information is used
11 for formative purposes. If you get 90
12 percent correct you move on because you've
13 achieved -- in a summative sense you've
14 achieved your goal so you move on.

15 But then if you don't achieve
16 your goal then they extract formative
17 information from those assessment results.
18 They look at what items you miss, what does
19 that tell you about the areas that you need
20 to study -- here you go, you go back and
21 study, you take the test again. So with the
22 same assessment both formative and summative
23 information is extracted from that.

1 And attitude -- the next thing
2 I'd like to invite you to think over -- to
3 rethink is that the attitude that
4 educational measurement is quantification.
5 Thorndike -- and I actually have this
6 written down because I always misquote it --
7 Thorndike said -- I'm going to misquote it
8 again because it's not here in front of
9 me -- that whatever is worth measuring can
10 be measured in some amount -- something like
11 that.

12 Well, I'd like you to rethink
13 that because that's an old idea. It's an
14 idea about -- that people learn by acquiring
15 more discrete facts or more discrete
16 approaches.

17 Bob Mislevy said in 1993 that
18 it's only a slight exaggeration to describe
19 current test theory as the application of
20 20th century statistics to 19th century
21 psychology. And that's the truth. With
22 item response theory, classical test theory
23 that's really the approach.

1 Modern theories of learning
2 portray the student as reorganizing,
3 restructuring, as their understanding
4 evolving, as becoming more complex, not as
5 acquiring discrete knowledge and facts.

6 So our assessment system should
7 show fidelity between our understanding of
8 how students learn and our method of
9 assessing it, so that if we describe
10 learning mathematics as acquiring
11 successively more complex understandings of
12 mathematics, we shouldn't then try to assess
13 it in a way that describes them as learning
14 discrete facts or algorithms.

15 The consequences of this lack of
16 fidelity may be no learning takes place
17 because the assessment information doesn't
18 fit the way teachers think about student
19 learning, but it can even be negative
20 student learning. What Dave Lohman has
21 talked about and I think Dick Snow before
22 him talked about is mathemafantic, actually
23 resulting in negative outcomes for the

1 student where they disliked math, they
2 disliked science, because of the way the
3 assessment information is provided back to
4 the teacher.

5 Next, rethink the definition of
6 an assessment. So right now we think of --
7 we often think of an assessment or a test as
8 a set of items in a test form. I'd like to
9 think -- like you to think about assessment
10 as a range of opportunities for the student
11 to show what they know and can do.

12 This is evidence in our design.
13 But they usually put it in terms that's even
14 less easy to understand than what I've just
15 said. But the idea -- and it's an idea that
16 extends across classroom assessment, it
17 extends across interim assessment or large-
18 scale assessment is that you are trying to
19 arrange the environment so that the student
20 has a chance to show what they know and can
21 do.

22 For example, here's a multiple
23 choice item, but it's arranged differently

1 than just scored -- keyed correct,
2 incorrect. Each of the four response
3 options is keyed to a different place on a
4 learning progression. So you can't just
5 score it right or wrong. If you score it
6 right or wrong you lack fidelity with a
7 learning progression underlying it. It's
8 arranged in a way to let the students show
9 where in the learning progression they are.

10 There's a computer based
11 interactive simulation. And what you get
12 from that is a path through a problem-
13 solving space. Again, if you score that
14 with an embedded multiple choice item during
15 the simulation again you're missing a lot of
16 information. You've arranged the
17 environment to show who a student problem
18 solves. You need to capitalize on the
19 information that you're collecting.

20 Finally, the attitude that -- am
21 I going backwards? Story of my life. So
22 we've covered the definition of an
23 assessment. So now I've invited you to

1 rethink some of those things let me talk a
2 little bit about a vision of a system for
3 learning -- not just an assessment system
4 but a system for learning.

5 And I'm going to introduce one,
6 two, three, four, five components. And each
7 of these components is a little different.
8 But in the middle you see the idea of a
9 learning progression -- sometimes called
10 learning trajectories.

11 So let me talk about that first.

12 A lot of our presenters up here have said
13 we need to know where we're going first.
14 They've talked about we need to know where
15 we need to get to before we can say how
16 we're going to assess it.

17 And that's exactly the same idea
18 here. Learning progressions describe how
19 students learn over time: In mathematics,
20 in science, in any of the subject areas,
21 they describe how students acquire
22 increasingly sophisticated understandings.

23 So you see the little graphic

1 there on the left. Learning progressions
2 can serve as a scale representing
3 achievement in growth for both classroom and
4 large-scale assessments. The idea is that
5 you can track how these increasingly complex
6 ideas and strategies and understanding
7 develop over time.

8 They can be broad; that is, they
9 can span years. They can be detailed and
10 only expand a few weeks of how students
11 learn a particular topic. And there are
12 examples of all kinds of those in the
13 research literature.

14 They can guide the types of
15 assessment tasks that fit different points
16 on the learning progression, because once
17 you know what sort of strategies and
18 understanding students are expected to
19 acquire, not what theta they're expected to
20 reach, it gives you an idea of how you can
21 arrange the environment to assess that.

22 They also -- sort of the other
23 side of that coin is you're supporting

1 pedagogical practices because you're looking
2 at what curricular tasks can we design, how
3 can we differentiate instruction between
4 those students who are at one point in the
5 learning progression versus another; how is
6 it that we can -- once we have in formative
7 sense identified a gap in students' learning
8 from where they are now and where we'd like
9 them to be, how can we then address that in
10 instruction. So the learning progression is
11 the glue that holds this system together.
12 It is what aligns it.

13 The large-scale assessment would
14 be computer delivered in this vision that I
15 have wherever possible. But it would be
16 objectively scored. When I say objectively,
17 I don't mean multiple choice scan, but I
18 mean mechanically.

19 So, for example, scoring using
20 artificial intelligence, like in thematic
21 analysis, for example, that's mechanical
22 scoring, and it's just as consistent as
23 scanning a multiple choice test. Having the

1 computer score an algorithm or follow a
2 problem space and compare that to the ideal
3 solution, that's also mechanical. Human
4 judgment's not involved at all. Again, it's
5 just as consistent as scoring a multiple
6 choice test.

7 So we'd have it objectively
8 scored. We'd have a variety of formats --
9 simulations, scenario-based tasks, multiple
10 choice questions, concept maps, activities.

11 The goal of having different formats for
12 the assessment is first, fidelity for what
13 it is you're trying to measure. The second
14 is efficiency.

15 For example, the multiple choice
16 test is probably an efficient way to measure
17 vocabulary. But it also matches the way
18 that vocabulary is often learned. On the
19 other hand a multiple choice test may not be
20 a good way to assess a student's
21 understanding of how to do an experiment --
22 that is, how do you arrange conditions in
23 order to do an experiment and observe

1 causation? Probably you should do that
2 either through a simulation or performance
3 based task.

4 The large-scale assessment would
5 also serve as an external moderator for the
6 classroom assessment I'll be talking about
7 in a little bit. So you could compare the
8 large-scale assessment result to the
9 classroom assessment results that are
10 aggregated up.

11 And if the classroom assessment
12 results don't match well -- let's say
13 they're at the -- if you did a regression --
14 let's say simple regression they were 5
15 percent in the group that matched -- in the
16 worst match and let's say 5 or 10 percent of
17 the worst match you might want to explore
18 why that was.

19 Maybe it's because the teachers
20 need more training in order to score their
21 classroom assessments -- they don't
22 understand the content as well or
23 something -- something like that. But they

1 are a more narrow representation of the
2 construct, and it's a one-time example of
3 student performance, so there that's
4 drawback.

5 But they are perceived as more
6 reliable and independent than the classroom
7 assessment. The classroom assessment would
8 be situated, teacher delivered, but may, of
9 course, not be teacher developed. That is,
10 you could have professional development
11 coming in and teaching teachers how to do a
12 clinical assessment, a clinical method, for
13 example, for assessing student
14 understanding.

15 Or you could have a bank -- that
16 I'll talk about in a little bit -- of tasks
17 that have been developed by other teachers
18 or by the state or by the federal
19 government. It would be used for diagnostic
20 but also summative purposes; again, include
21 a variety of formats, but instead of
22 simulation-based and scenario-based
23 computer-delivered tasks, you would have

1 performance-based tasks.

2 You would ask teachers as the
3 primary assessors, but as a number of the
4 presenters up here have said -- I'm echoing
5 this -- you need to have sustained
6 professional development to do this well.
7 You can potentially assess broader aspects
8 of student understanding and you can assess
9 it over time.

10 It's not independable; it is
11 perceived to be more susceptible to
12 inconsistency and bias. The challenge --
13 also there's a challenge: How do we record
14 sufficient information from the classroom in
15 order to make decisions about students? Can
16 we take advantage of technology? For
17 example, my server the other day in the
18 restaurant brought to the table a little
19 device where they ran the credit card.
20 Wireless -- right? -- wireless -- the
21 entered information on it just like that.
22 Why can't our teachers have a similar device
23 to enter in student information? That kind

1 of technology already exists.

2 Let's see. Online support
3 bank -- the fourth component would be an
4 online support bank. This would support the
5 everyday practices of teachers. It would
6 include guidelines for instructional and
7 assessment activities.

8 Learning examples -- it's very
9 important to show teachers learning examples
10 that are annotated about what does this show
11 about students' status on a learning -- on
12 the learning progression. Video formats --
13 because there's nothing more powerful than
14 actually watching students perform.

15 There would be professional
16 resources, such as articles or presentations
17 from other teachers or other professionals.

18 An online forum for informal teacher
19 collaboration and discussion. Again, these
20 kind of banks already exist.

21 Finally, data capability -- and
22 this is very important because if you're
23 going to use classroom assessment as well as

1 a large-scale assessment as a summative
2 assessment you have to have a way to
3 collect, aggregate, and interpret data
4 across observations, time, students,
5 classrooms, and perhaps even states. So you
6 have to have a common data format. And
7 that's critical to aggregation.

8 You also have to have a common
9 understanding of learning because if you
10 don't have a common understanding of
11 learning across classrooms and time and
12 states then how can you interpret the
13 results if -- how can you interpret the
14 student performance?

15 Also the data capture must not be
16 intrusive for classroom assessment. You
17 know, that teacher/student interaction has
18 to be preserved, and so the way to capture
19 the data has to be done in as least
20 intrusive and most efficient way as possible
21 to allow the teacher to teach.

22 I suggest a 50/50 combination of
23 classroom and large-scale assessment. Carol

1 talked -- gave some examples of other -- at
2 the international level of this being done.

3 There are ways out there -- proven ways to
4 combine classroom and large-scale assessment
5 to do summative evaluation.

6 One of the important things to
7 note about learning progressions is the
8 information is ordinal in nature. Right?
9 It's not interval. It know this is sort of
10 a psychometric thing. Interval information
11 says that learning is done in a way that you
12 can divide it. Right? You can divide this
13 student's learning by that student's
14 learning.

15 But that's not how we understand
16 student learning. We understand student
17 learning in terms of qualitative terms. And
18 so the scale that we use to measure it
19 should also be qualitative and ordinal.
20 That's more easily consumed by teachers,
21 students, and parents and policy makers as
22 they try to make decisions to increase
23 student learning.

1 Is this practical? Well, the
2 system of coordinated assessment and
3 instruction I have described is more than a
4 vision. It actually exists in one form or
5 another out there in different places.

6 If you look at the Gismo Research
7 Laboratory at North Carolina State
8 University working with the state of North
9 Carolina they've developed a set of content
10 standards that incorporate learning
11 trajectories and learning progressions.
12 They've developed an online database.
13 They've developed a set of what they call
14 Gismos -- and you can find all of these on
15 the web -- at their web site.

16 And it all -- it's all researched
17 based. It's all practical. They've done a
18 field test with I think about 5,000 students
19 that they're currently scoring across grades
20 K through eight I believe.

21 Another example is a system from
22 Queensland, Australia. And you can read
23 about that in a paper by Stanley, Gardner,

1 Reynolds and Wild. Stanley is a chair at
2 Oxford, and I think you can find that paper
3 online as well. It describes the systems in
4 Australia, in Scotland, in Wales, and
5 England. It's a very good summary of all
6 those different systems.

7 So this is not just an approach
8 that is possible ten years from now. It's
9 an approach that -- whose components are
10 being used now.

11 Technical qualities of this
12 vision -- what is a reliable system for
13 learning? Remember, we're focusing on a
14 system here -- not a test. So a reliable
15 test doesn't mean that you consistently get
16 positive learning outcomes.

17 What we're looking for is a
18 system that consistently produces positive
19 learning across classrooms, across students
20 from different ethnic backgrounds, different
21 cohorts in time. That's a reliable system
22 for learning.

23 In terms of validity we're

1 stressing consequences -- the consequences
2 of using test score information -- or
3 assessment information. That's the
4 important thing here. It's not -- doesn't
5 correlate with something, is not content --
6 or content is important -- but it's do we
7 see positive consequences. And we don't
8 often hear consequences talked about when
9 we're talking about the validity of test
10 scores.

11 And we have to make clear the
12 components that lead to these learning
13 games. That's the important thing. What is
14 causing learning games in the system.

15 Finally, how do we create
16 vertical alignment on the entire system? We
17 do this using learning progressions. Let me
18 go back for a moment. What we have are
19 learning progressions. The understanding of
20 how students learn and perform is what
21 unites all these different components -- the
22 online support bank, the large-scale
23 assessment, data aggregation/interpretation,

1 classroom assessment. It's all tied
2 together by our understanding of how
3 students learn and perform over time.

4 And that provides the common
5 scale, as it were, that we psychometricians
6 like to talk about -- the common scale.
7 That provides a thing that ties it all
8 together and allows information to be passed
9 back and forth, allows students to be tested
10 at different points in time, allows data to
11 be aggregated, allows classroom assessment
12 to fit with the assessment information
13 coming from the large-scale assessment.

14 And that's all I have. So thank
15 you.

16 MS. WEISS: Thanks. That was
17 terrific. Questions?

18 MS. WURTZEL: Thanks so much,
19 Paul. So you noted that all the components
20 of this system that you sketched out are in
21 existence now. So I have two questions,
22 which is which are the biggest stretch to go
23 to the scale, and if all the components

1 exist now what have been the reasons why
2 states have not embraced them? What have
3 been the barriers or the decisions that
4 have -- what people think that wasn't the
5 right way to go.?

6 MR. NICHOLS: Yeah. One of the
7 biggest stretches is actually data
8 management. We've been trying to put
9 together a system in another professional
10 area where we're trying to coordinate
11 diagnostic information, summative
12 information, and instruction and in delivery
13 of materials.

14 And just trying to coordinate
15 across platform is just a headache. I think
16 that's one of the biggest stretches right
17 now -- coordinating across platforms --
18 which I know nothing about, but I just
19 observed it.

20 One of the reasons I think it
21 hasn't been done is I believe that -- and
22 this is a stretch. But I think that
23 conventional test theory really does stand

1 in the way of some of these things. Once
2 you throw IRT on top of something you start
3 distorting the information that comes from
4 it.

5 The idea that you have to have
6 that reliability as this narrow sort of idea
7 that also constrains what you want to do.
8 For example, I've seen in other countries
9 where they give scores to essays through
10 social moderation. We can't do that. We
11 have to have independent assignment of
12 scores so that we can assess the reliability
13 of of writers. Right? We can't assess
14 agreement across writers who socially
15 moderate the result. So we have independent
16 scoring. So those sorts of --

17 MS. WEISS: What does socially
18 moderate mean?

19 MR. NICHOLS: They discuss and
20 assign a score. So they'll sit there and
21 say, Well, what did you like about this and
22 what was the strength, what did you see, and
23 then they'll assign a score.

1 But if you look at independent
2 observations in a statistical model, you've
3 got correlated errors doing that and you
4 can't get a reliability estimate. So those
5 sorts of ideas that we really are pulling
6 from and experimental psychology of the 19th
7 century really do get in the way.

8 MR. POSNY: It was an interesting
9 idea when you talked about a 50/50
10 combination of using the -- you know, the
11 classroom assessment and that large-scale
12 assessment as well. What's your -- you
13 know, why 50/50? And could there be, you
14 know, a different amount depending upon the
15 needs or the particular child?

16 MR. NICHOLS: I think there could
17 be. I don't think there's any magic in
18 50/50. I sort of -- it's been -- it's used
19 in other countries. I've seen it used in
20 other countries and so it seemed like a
21 reasonable thing to do. But there's no
22 magic in 50/50.

23 And I don't see -- what you want

1 to do is to create an accurate understanding
2 of what that student knows and can do, so
3 50/50 can certainly be varied to get what
4 you might call construct validity.

5 MS. WEISS: Can you talk a little
6 bit more about the large assessment? What
7 I'm wondering is what are the attributes of
8 that assessment that are different than what
9 we're currently doing now? Or do the
10 current assessments -- are the current
11 assessments good enough to be that large-
12 scale assessment, and we should focus in the
13 other places?

14 MR. NICHOLS: It's different in a
15 number of ways. One of the things that
16 we're -- one of the projects that I work on
17 is a project applying evidence-centered
18 design to the scenario-based computer-
19 delivered Minnesota Science Assessment.

20 And that's different from a
21 computer-delivered -- a routine computer-
22 delivered assessment. First, it's scenario-
23 based, so when you develop it you develop it

1 in a whole new way. And a simulation you'd
2 also have to develop it an entirely new way.

3 They actually developed that like you might
4 develop a comic book or a movie in that you
5 sketch it out frame by frame how the
6 scenario develops.

7 And so that's why we're trying to
8 use evidence-centered design, because right
9 now it's a very labor intensive way of
10 developing that and we hope through
11 evidence-centered design to lessen the cost
12 and lessen the time it takes to develop
13 that.

14 The other thing is that you're
15 using -- for example, there's a set of
16 characters that we use. It's a set of
17 characters that the state has approved --
18 just almost like a set of terms. So you've
19 got these characters.

20 MS. WEISS: It's a new law --
21 state approved characters?

22 MR. NICHOLS: Yes. And the first
23 set they came up with looked a bit like

1 zombies -- truly they did. It was like
2 Night of the Living Dead in the science
3 assessment. So they came up with a new set
4 that had more life to them -- that were more
5 attractive -- so those sorts of things that
6 are part of test development that you never
7 had to consider before.

8 The other thing is that you've
9 got -- it takes the passage-based idea --
10 and, again, this is getting into the
11 psychometric weeds a bit -- but it takes the
12 idea of passage-based assessments to a whole
13 new level because now you have all these
14 observations, again, are not independent.
15 So if you're working inside a simulation --
16 not independent at all.

17 It's very hard to break out of
18 the mind set of a test form as a set of
19 items, whether they're computer delivered or
20 not -- as just a set of items that are
21 scored right or wrong because that's not how
22 you're scoring these things.

23 And so that challenges

1 everything. How do you equate? How do you
2 get item statistics? How do you --
3 everything is challenged. It's not
4 something we can't overcome -- it's just
5 something we have to think about.

6 MS. WURTZEL: Paul, you made a
7 point that some of the experts raised in our
8 last meeting in Boston about the balance
9 between reliability and validity and talked
10 about how to think about consequential
11 validity in the context of assessments that
12 are tied to college and career readiness.
13 Could you elaborate that a little bit more?

14 MR. NICHOLS: College and career
15 readiness, reliability and validity. Okay.

16 First of all, I have said before you have
17 to have validity before you have
18 reliability. That is, you have to
19 understand what it is you're assessing
20 before you can understand whether you're --
21 when it is good to estimate consistency
22 across what you're assessing. So that's the
23 first thing -- and it is reliability and

1 validity intertwined and you can't pull them
2 apart.

3 In terms of college readiness,
4 again, we don't really know what that is.
5 It's a very vague term. But if you're
6 thinking about some of the things that Carol
7 talked about -- problem solving, critical
8 thinking -- all sorts of things -- again,
9 the concept of reliability in terms of
10 consistency across situations in which you
11 think performance should be consistent --
12 that's sort of a broad definition of
13 reliability.

14 That's applies just as well to
15 assessing critical thinking as anything
16 else. There's really no -- there's no head
17 butting between validity and reliability.
18 If you're going to get consistency across
19 where you think there should be consistency
20 it's validity evidence as well as
21 reliability evidence in that case because
22 it's confirming your -- it's evidence that
23 tends to confirm your understanding of what

1 it is you're assessing.

2 MS. WEISS: Other questions?

3 Okay. Maybe we should do the Brian Gong
4 two-minute stretch, And then we are going to
5 launch into our roundtable discussion

6 (Whereupon, a short recess was taken.)

7 MS. WEISS: Okay. So now we are
8 going to have a discussion up there. Thank
9 you. I've gotten a number of comments and
10 question from folks out there. Some of them
11 we've hit on a little bit in some tangential
12 ways.

13 But the first thing that I'd like
14 to do is just spend a couple of minutes -- I
15 think that I heard some consistency around
16 the question of how many assessments are
17 there in an assessment system and how does
18 that system work. And I just want to sort
19 of see quickly where we are on the same page
20 and where we're not.

21 Maybe we can start with the
22 freshest one in our minds. Paul's had
23 basically two big picture components

1 supported by some professional development
2 and other support information, but sort of a
3 classroom based component that was
4 throughout the year and large-scale
5 component that was end of year and somehow
6 the two of them in some balance -- like
7 50/50ish -- together equal a summative
8 score.

9 Let me just have you guys, if
10 you've got a different picture in your head,
11 just say quickly where the picture differs
12 from that one and that will just help us I
13 think anchor in some of the rest of the
14 questions that we're going to get into after
15 this.

16 MR. FABRIZIO: I guess one of the
17 comments or questions that I would ask Paul
18 and maybe everyone else is, you know, we
19 don't really have states that are doing that
20 combination where -- it's other countries.
21 And is it the fact that in these other
22 countries people believe what teachers say
23 or -- I mean, what -- why is it that we have

1 what I guess is a pessimistic view -- and
2 Tom talked about I guess grade inflation
3 or something.

4 And this was a question I thought
5 about earlier that I didn't get a chance to
6 ask you, Carol. And when you say the
7 teachers scored, are those teachers scoring
8 the actual work of the students they are
9 teaching or do you at least make sure that
10 teachers aren't scoring their own students'
11 work.

12 MS. CAMPBELL: So on that
13 specific question, Lou, there's multiple
14 layers to it -- and I might not have
15 explained it fully at the time. On teacher
16 moderation on summative assessments -- so
17 the large-scale tests -- in Ontario that is
18 grade three, six, nine, and ten -- not any
19 other years -- so four years out of the
20 student's 12 years or whatever education.

21 Those tests are just under 50
22 percent multiple choice, just slightly over
23 50 percent open-ended constructed response.

1 And teachers and other examiners will be
2 involved in scoring those tests.

3 The teachers self-select -- it's
4 work over the summer, it's paid work, it's
5 trained work. They wouldn't be testing
6 their own students. It would be blind and
7 it would be peer refereed and all of those
8 types of processes -- and similar in other
9 parts of Canada.

10 The pieces where the teachers are
11 involved in scoring their own students' work
12 is the ongoing course work and the extended
13 projects. On those in England -- because in
14 England they do now count in terms of end of
15 year marks -- the balance there is that the
16 course work or the project, the controlled
17 assessment as it's called, is determined by
18 external criteria and regulatory
19 qualifications. So it's within a criteria
20 and framework that the teachers would then
21 assess. Or they can be locally developed
22 but then they have to be externally
23 assessed. There's a process of checks and

1 balances in all of the pieces.

2 MR. NICHOLS: I would just add
3 there is a particular American context for
4 assessment. I think it's going to be a
5 public television series in context.

6 (General laughter.)

7 MR. NICHOLS: But -- it really
8 is. It's developed and there's a
9 particular -- there's this technology of
10 validity and reliability and multiple choice
11 tests that I don't see in other countries.

12 MR. GONG: Quick response. I
13 think that there was agreement about a state
14 and a local component. I think there was a
15 wide disagreement about what function the
16 local was doing and what the design was.
17 And I think that -- I think you have to
18 have -- decide what the purpose is before
19 you say what the local one is.

20 So I'm in favor of having both.
21 I think it's something that can be done. I
22 think it's something that would take -- this
23 is sort of my second generation. But my

1 main concern is because of both validity and
2 reliability. I think that -- the short
3 tests we have are not valid in that they
4 don't sample enough of the things that we're
5 really interested in. And I don't think
6 they're -- they're reliable within
7 themselves, but they're not reliable across
8 the application thing.

9 So if you ask a student can you
10 do all these things you would know that
11 better if you had a wider sample of types of
12 evidence. So I'm in favor of doing that.

13 But in terms of local, we would
14 have local -- so here's -- you could local
15 decides on the test specifications, they
16 decide on the goals, they decide on the
17 design of the test, they decide on the
18 development, the scoring.

19 And I think that some of the
20 systems we're talking about actually all of
21 those and some had very few. Some were
22 things that the state controlled almost all
23 of those and teachers were implementing

1 them. I think that when we talk about state
2 versus local we ought to clarify how local
3 we're talking about. But I agree everyone
4 mentioned it. I think we were probably
5 talking about different things.

6 MS. WEISS: Yeah. Go ahead.

7 MR. FISHER: This last few
8 minutes of conversation gets back to my
9 point that I made about Algebra I and
10 consistency. If we really believe that
11 we're going to adopt common academic
12 standards and expectations, whether it be a
13 high course or for third grade reading or
14 whatnot, then it makes sense that teachers
15 will be trained, in-service particularly --
16 maybe pre-service -- to teach to those
17 things that we have adopted as content
18 expectations.

19 I mean, that's what we want them
20 to do. We don't want them to go wandering
21 off doing the volcano projects simply
22 because the kids like the volcano project if
23 it's not related to what it is they're

1 supposed to be learning. And it's fair game
2 for all the sixth grades teachers in
3 mathematics to have the same expectation for
4 what they want kids to know when they leave
5 sixth grade.

6 So I don't have any trouble at
7 all with the concept that there's a huge
8 need for training of teachers to do a better
9 job and a more consistent job and lead kids
10 toward higher expectations. I separate that
11 in my mind from the issue of what a state or
12 national government may want to have for
13 data collection for other high stakes
14 purposes. Not everybody agrees on the
15 concept of high stakes for high school
16 graduation. We already said that -- a
17 couple of people have said that -- there's
18 only half the states that follow that.

19 But I think 100 percent of the
20 states want data to make more informed
21 decisions about progress, about allocation
22 of resources, about understanding
23 educational phenomena. All of these things

1 are things that you can do when you crunch
2 numbers. So you can have a state assessment
3 program that has no high stakes but it
4 generates numbers, or you can have one that
5 generates numbers and has high stakes.

6 MR. GONG: Follow-up question.

7 Tom, I agree with you that we need to
8 have -- on the Algebra I it's an interesting
9 example -- that we need to better specify
10 what it is that we really expect.

11 Could you talk about the
12 relationship between that type of
13 delineation and curriculum? Now, let me set
14 it up a little bit. The NGA CCSSO end of
15 high school draft common core state
16 standards in math have concepts and skills.

17 There are 50 concepts and 33 skill
18 standards. So if you put those in detail
19 you have about one per week of a school
20 year. Okay.

21 So if we -- now we're going to
22 put more detail on them so that we
23 understand what the level of expertise and

1 proficiency and independents and things are.

2 So it sounds like we're getting close to a
3 curriculum. Would you -- but maybe we're
4 not entire there. Could you talk about
5 what's needed for your assessment and how
6 that's different from a curriculum?

7 MS. FISHER: Well, I'm certainly
8 not an expert on what they're proposing as a
9 draft. But I was a math teacher for eight
10 years, married to a math teacher. And my
11 doctorate's in curriculum, believe it or
12 not.

13 You can have statements of
14 outcomes such as the ones they're
15 development but you still have to turn it
16 into a sequences curriculum and
17 instructional program for delivery to the
18 students. And textbooks publishers are
19 still going to be selling textbooks, and
20 they've got to model it after something.

21 So I have not read anything that
22 says that NGA CCSSO is going to be in the
23 textbook publishing business, so I assume

1 that once those things are out there
2 somebody's got to do that. Somebody's going
3 to have to turn it into an operational
4 curriculum that will be delivered to
5 students. And, in addition to that,
6 somebody's going to have to translate it
7 into test specifications and specific sub-
8 objections, which has not been done.

9 So I sort of view what they've
10 done more at the goal level than, you know,
11 at the day-by-day and week-by-week
12 instructional level. I haven't heard any
13 conversation about how they propose that
14 that be done. Perhaps you've heard some
15 gossip about it; I have not.

16 MS. WEISS: Thanks. I mean, I
17 think you're right, Brian. I think that by
18 starting just with Paul's as an anchor there
19 was -- there were other points of view,
20 particularly around the classroom based part
21 and who developed it and what the purposes
22 were and how many of them and how
23 frequently. And so I do think there was

1 some consensus around the fact that there's
2 these two parts, but both parts need more
3 fleshing out and thinking.

4 Which maybe leads me to a
5 question that we've been struggling with a
6 lot around -- let me just sort of put out
7 there a chain of sort of two or three
8 questions that feel interrelated to us even
9 though they sound a little different.

10 So how many consortia do we fund,
11 which is related to how many tests do we
12 need, which is related to where do we need
13 the innovation to happen because we don't
14 know the right answers so we need multiple
15 models and multiple examples? And where do
16 we know something and ought to be tighter in
17 specifying, You know, what, we only need one
18 of these but we need many of these?

19 So I know this is a giant
20 question, but let me throw it out there and
21 then maybe through additional probes and
22 things we'll hone it down a bit. But --
23 yeah, go ahead.

1 MR. NICHOLS: I have a question
2 first. Are we assuming --

3 MS. WEISS: No, no, no. You're
4 not allowed.

5 MR. NICHOLS: That what you get.
6 When you work with psychometricians, you
7 ask a question you also leave with three
8 more questions than you had.

9 Are we assuming that there's a
10 common understanding of what students are --
11 should learn and be able to do before we
12 start developing these different
13 consortiums?

14 MS. WEISS: Yes. So you're
15 asking whether all the consortia have the
16 same common standards? Well, for the
17 purposes of this conversation, let's say yes
18 to that and just narrow one variable a bit.

19 Yeah, go ahead, Teri.

20 MS. SISKIND: I'll start with
21 just part of it. I mean, I think we need --
22 I think that some of us from the states do
23 have the directive through the NGA and

1 chiefs' compact that there is the set of
2 standards for which our leadership would
3 want a common assessment. And I don't
4 know --

5 MS. WEISS: Right.

6 MS. SISKIND: -- whether you want
7 to take a show of hands on that. But I
8 think that that would be -- I think for some
9 of us having at least that one consortium
10 would be of import to our states.

11 MS. WEISS: And I guess what I'm
12 asking is, assuming we have one set of
13 common standards, is there a benefit to
14 having multiple assessments to test the same
15 standards or not -- or in what cases or at
16 what grade levels or subjects or --

17 MS. SISKIND: Well, I've already
18 expressed --

19 MS. WEISS: Yeah, you --

20 MS. SISKIND: -- you know, my
21 point of view on that. And I think, you
22 know, if -- again, if our purpose is --

23 MS. WEISS: Common -- be common.

1 MS. SISKIND: Right. And common
2 in a good sense.

3 MS. WEISS: Yes.

4 MS. SISKIND: I do think that
5 there are other consortia that many of us
6 would be interested in for work that we
7 don't feel most comfortable with for
8 advancements, for example, with a 2 percent
9 test or something like that. And I'm not
10 sure whether that would be one consortium or
11 two.

12 You know, I think with both of
13 these -- back to something Lou and I I think
14 both said -- is we'd want some exemption
15 from duplicate testing --

16 MS. WEISS: Yes.

17 MS. SISKIND: -- and some sort of
18 guarantee that we'd be shepherded in the
19 right direction and that if we went to peer
20 review we wouldn't -- we would experience
21 some success.

22 MS. WEISS: I heard that part.

23 MS. SISKIND: So I think -- I

1 mean, I've got other components too but I'll
2 stop there. I think --

3 MS. WEISS: What do others think
4 about whether it's --

5 MR. FABRIZIO: I think it needs
6 to be given that whatever assessments or
7 systems are developed during this creative
8 period of our education measurement time
9 line that they should be based on common
10 core standards. I think that should be a
11 given.

12 I mean, everything that I have
13 been hearing and -- you know, and I've said
14 it publicly, I am so tired of having to
15 defend why our results are the way they are
16 and why it's different than South Carolina
17 and why it's different than Virginia. I
18 mean, my life would be a heck of a lot
19 easier to be able to say the consortia of
20 states all got together and this is what
21 we -- and doesn't matter which state you
22 live in, this is what -- we've all agreed
23 kids need to learn.

1 So my quick answer would be you
2 should have multiple consortia all trying to
3 come up with different ideas. To me this is
4 an -- the federal government right now is
5 coming up with money to help fund creative
6 things. And if you put too many constraints
7 on it you've lost what I thought the whole
8 purpose of this was for.

9 And so I'd like to see different
10 consortia come up with different types of
11 systems. And then at the end we -- you --
12 evaluate what are the pluses and minuses of
13 these different systems. And then you go to
14 the next step after that. I mean, that's
15 the only thing that made sense to me.

16 MS. WEISS: So you would let
17 different consortia do different --

18 MR. FABRIZIO: Different things,
19 yeah.

20 MS. WEISS: -- tests against the
21 common -- against --

22 MR. FABRIZIO: As long as they're
23 all --

1 MS. WEISS: -- a common set of
2 standards.

3 MR. FABRIZIO: That's right. And
4 then we check to see how well is that
5 working.

6 MR. NICHOLS: I would add to what
7 Lou said. And I think there needs to be
8 clear criteria established. There should be
9 articulation between research and practice.

10 You would require empirical evidence of the
11 consequences of these systems so that you
12 can evaluate one against another. And it
13 has to be built in up front about what you
14 expect that to happen -- and you expect
15 these sorts of criteria.

16 MS. WEISS: Yeah, go ahead,
17 Brian.

18 MR. GONG: I agree but for a
19 different reason. States don't typically
20 when they let an RFP for a state assessment
21 have competing ones. And that may say why
22 they don't have much innovation.

23 But if you know what you want

1 then issue one RFP, give one award, and make
2 sure that it's delivered right. But if you
3 don't know what you want then you need more.

4 And if you aren't sure that you're going to
5 get it then you may -- then you need to do
6 multiple ones.

7 I think there are some things,
8 especially on these near term ones, where
9 you actually know them. But on high
10 school -- so this is an example of what you
11 want. High school you might say, I want an
12 end of course and a survey. And the reason
13 you might want that is that's sort of
14 current practice.

15 But I would ask what you -- what
16 is it that you really want to know. Do you
17 want to know whether students retain the
18 information and can apply it -- not
19 immediately after they took the course but
20 some time after? Then I would ask for a
21 survey thing that does application some time
22 after the course.

23 If you want to know something in

1 detail about those specific things then an
2 end of course is better. Start with what
3 you want to know, not this convenience
4 about -- just about the administration.

5 So I -- but I think you have end
6 of course and survey. I think you have a
7 multiple choice -- and a multiple choice and
8 constructed response. You cross those --
9 now you've got four combinations. And I
10 think you've got a computer based and a
11 paper and pencil. Now you've got six.

12 And I think you have to decide
13 whether you want to fund some of those, all
14 of them, or -- and this gets back to the
15 incentive. You may say, You know, I've
16 already got all six of those running. What
17 is it that I really want out of this? Why
18 don't -- am I getting something different?

19 Then I could go to these states
20 and say, This state has an end of course,
21 multiple choice, paper and pencil -- I'll
22 bless it or something and people can do that
23 one or -- so, anyway, that's a simple

1 example of high school where you could end
2 up with six variations. And I don't think
3 you ought to fund them all. I think you
4 ought to sponsor them in some way, but I
5 don't think you have to fund them.

6 So I think we could work through
7 all these other examples and -- you're going
8 to have more permutations of these than you
9 can fund. So, again, I think it comes back
10 to what --

11 MS. WEISS: Right. And,
12 remember, we're developing a competition
13 here.

14 MR. GONG: That's right.

15 MS. WEISS: So we're going to
16 fund as many that has ranked very high
17 quality until the money runs out.

18 MR. GONG: Yeah.

19 MS. WEISS: So it's not about
20 funding them all.

21 MR. GONG: That's right.

22 MS. WEISS: It's about what are
23 the characteristics of high quality --

1 MR. GONG: Right.

2 MS. WEISS: -- makes something
3 worth funding --

4 MR. GONG: Right.

5 MS. WEISS: -- is sort of what
6 we're trying to --

7 MR. GONG: And I would really --

8 MS. WEISS: -- figure out.

9 MR. GONG: -- really distinguish
10 between using money to get states to
11 participate versus getting some product at
12 the end that's portable. Because I think a
13 lot of the way states are viewing this is if
14 you give me money then I'll participate.
15 And I don't think that's --

16 MS. WEISS: Right.

17 MR. GONG: -- that's -- you won't
18 have enough money to make that work. I
19 think you're trying to develop products that
20 states will find attractive so that they can
21 adopt them.

22 Just as a note of -- we're
23 assuming about the common core standards, if

1 you went to states and said, I -- how many
2 of you believe that these are both college
3 and work ready and give me your test
4 specification for what the relative emphasis
5 will be on college and work ready. Some
6 people will say they're exactly the same and
7 some people will say they're not overlapping
8 very much.

9 I would not assume that the CCSSO
10 NGA standards are going to give us a unity
11 without doing what Tom said -- you have to
12 go through the test specifications in order
13 to really find out how close they are -- at
14 least test specifications, as well as use.

15 Just a last note -- I want to go
16 back to use -- about how many of these you
17 really want. I'm going to ask what you want
18 to use them for because there are some uses
19 where two is really a bad idea -- it's one
20 too many. And there are some cases where is
21 pretty. You know, some say, Well, I want a
22 college admissions test. Well, you've got
23 SAT and ACT and we've done pretty well with

1 two. We don't need to collapse that market
2 into one. So I would go back to the use --
3 what is it that you really want to have
4 happen.

5 MS. WEISS: And the other
6 question we're wrestling with is what -- so
7 it's also an innovation question. In places
8 where we don't know a lot do we foster
9 innovation and couple it with evaluations so
10 that we can learn something.

11 And so that's partly what I think
12 we're asking you -- not how many is the
13 right number, but where are the places where
14 need innovation because we as an industry
15 don't know the right answers, and where we
16 do know the right answers then we should
17 just go for it.

18 That's a distinction that -- you
19 know, when you say it out loud it sounds
20 like a clear distinction, but we know in
21 this field it's not clear what the answer is
22 to that. So that's partly I think what
23 we're trying to ask you for expert thoughts

1 on. Go ahead, Carol.

2 MS. CAMPBELL: I think to build
3 on what others have said around the
4 consortia piece and the common standards, I
5 mean, I'd like to echo some of what's been
6 said. It strikes me if there is agreement
7 on common standards there needs to be some
8 degree of commonality around assessment
9 criteria. And that can be at various levels
10 with specificity.

11 But -- to have common standards
12 but not a common understanding of what the
13 types of assessments for them might be I
14 think could be problematic. You know, I can
15 certainly talk about England and Canada.
16 Let's use England as the example. There are
17 national criteria, a national frame work,
18 and national specifications, but they're
19 fairly broad. But they set out this sort of
20 expectations for how standards will be
21 reached.

22 Then there's five exam boards
23 through a competitive process that they

1 actual assessment design and the item
2 constructs and all of that vary and schools
3 and districts can choose between them.
4 They're available online so you can actually
5 get access to the test items after they've
6 been used -- so they're sharing.

7 So I do think part of the
8 innovation is we -- if this is a pilot phase
9 two actually make what the consortia are
10 doing available across consortia and may be
11 after a period of I don't know -- two,
12 three, four years -- begin to look at sort
13 of best practices within and between.

14 The other piece I'm wondering as
15 you're talking to one another is whether the
16 expectation in the RFP is the old consortia
17 bid for all assessments or whether it is
18 that part of the innovation is that there
19 are particular consortia that are strong on
20 elementary or high school or on ELL or
21 special education rather -- I mean, I don't
22 know because I'm not --

23 MS. WEISS: Yeah. That is

1 actually one --

2 MS. CAMPBELL: -- a U.S. expert.

3 MS. WEISS: -- of the questions we
4 have -- is there a way to take apart these
5 things so that a consortium is specializing
6 in one area or another, or does doing that
7 make us lose the vertical alignment. So
8 what do you think of that? Where might it
9 make sense for a consortium to specialize?
10 And where would you recommend that we not
11 go?

12 MR. GONG: I think what Carol
13 said is really important. If there's
14 agreement on some -- at some level it helps.
15 And so if you can -- if the U.S. Department
16 of Ed can specify general assessment
17 parameters or even more specific ones then
18 it ensures that there's more coherence
19 across it. To the extent that there isn't
20 then then it will be by -- you can think
21 about the mechanisms to help make it happen,
22 but it's not as sure.

23 But I think that breaking them up

1 is a really good idea. I think there are
2 different interests and there are different
3 competencies that you want to have.

4 I want to come back to this idea
5 about why -- what the states' role is and
6 why there is a consortia. Because if you're
7 trying to develop a product what you're
8 doing is you need to have some competence
9 and some commitment to it.

10 And so I can think of some
11 reasons why multiple states being involved
12 in that is important, but if the states are
13 going to hire a contractor to do it then I
14 think you're actually getting a contractor
15 not -- you're -- and when someone talked
16 about states developing items I don't know
17 any state in the Department of Ed that
18 develops items.

19 They've got two math consultants.

20 They're hiring someone else. Whether it's
21 teachers or whether it's another contractor,
22 most Departments of Ed are not -- state
23 Departments of Ed are not doing it. So they

1 are going to be involving some other support
2 mechanism.

3 So look at that all the way
4 through and then think, Why am I having the
5 states in here. I think many of them
6 thought, I needed the money because this is
7 the thing I'm going to adopt. If the model
8 is the consortium you're in is what you're
9 going to adopt then these breaking it apart
10 is a bad idea.

11 But if the idea that -- is you're
12 developing products that people are going to
13 adopt crosswise then there's a very
14 different model for why states are involved.

15 MS. WEISS: So what if the model
16 is the second -- that there's different
17 products that will be adopted across these
18 consortia? Then how would you break it up?

19 Because it strikes me that still might be
20 bad ways to break this up and smart ways
21 that it could be broken up. In other words
22 would you -- is it a good idea to separate
23 elementary from high school assessments?

1 MR. GONG: In general, no. But
2 there are ways to get articulation for that
3 in order to do that.

4 MS. WEISS: That's right.

5 MR. GONG: I think states would
6 do that. States will divide up a high
7 school team and a elementary team and
8 they'll create ways to make them talk back
9 and forth.

10 MS. WURTZEL: So can I follow up
11 from a different direction, which is I have
12 lots of stars in my notes next to Carol's
13 comment about a common framework around
14 assessment.

15 MS. WEISS: Me, too.

16 MS. WURTZEL: And, Brian, you
17 picked up on that. And so could you flesh
18 that out a little more? What would that
19 look like? What are some examples of
20 criteria and then other people's thoughts
21 about that? Because obviously if you are
22 going to break this up having some common
23 framework helps when you try to put them

1 together at the end.

2 MS. CAMPBELL: Yeah. So the
3 framework or the criteria would ideally flow
4 from the standards -- or we're talking
5 having not seen the standards. But they
6 need to be aligned with standards and
7 curriculum and all of those pieces.

8 And then I guess it would also be
9 broadly around the principles -- and the
10 Department's already set out a range of
11 principles. I mean, I've got some in front
12 of me actually because I mentioned some in
13 my presentation, albeit quickly. So these
14 are some of the criteria in England for
15 functional skills, which are practical.
16 There's also criteria for the subject areas
17 in these ones.

18 But -- and I said the assessment
19 must provide realistic context, scenarios,
20 and problems, specify tasks that are
21 relevant to the context, require application
22 of knowledge, require problem solving.

23 So you can see they are fairly

1 specific but they're not at the level of --
2 this is a three-hour test with 30
3 constructed items. But it sets out the idea
4 of problem solving or it could be subject
5 knowledge. And there's lots and lots of
6 documentation around that that I can
7 certainly make available.

8 But it's that sort of -- it's a
9 sort of mid-level -- it's between the
10 standard and it's between the actual test
11 construction, but we expect all tests to
12 have these dimensions to them.

13 MR. NICHOLS: I can add to that.

14 First, the idea of a product -- I mean, I
15 think what you should be looking at is
16 methods and processes. It's sort of the
17 difference between giving someone fish
18 versus teaching them to fish. If you're
19 just focusing on delivering products at the
20 short term it will work for the next five
21 years.

22 But understanding what assessment
23 processes are going to give you -- satisfy

1 those criteria that you're looking for,
2 helping students learn, developing teachers,
3 and all of those -- that is where you're
4 building your knowledge base and where your
5 different projects can help.

6 So, yes, they will provide a
7 product, but the focus should be on building
8 your knowledge base around what assessments
9 and what aspects of those assessment
10 instruction systems are going to produce
11 what you're looking for.

12 MS. WEISS: Right.

13 MR. NICHOLS: That's why you need
14 to build your knowledge base -- not
15 products. That's --

16 MS. WEISS: Right.

17 MR. NICHOLS: That's the fish.
18 And that's part of that -- I think what
19 Carol is talking about is part of that
20 knowledge base that you're building when
21 you're doing that. And how can you assess
22 those things?

23 MS. WEISS: Right. And I think

1 product wasn't really necessarily like
2 something in a box as much as it was the way
3 Brian was talking about -- a model that's
4 different from everyone in a consortium buys
5 into every test in that consortium and
6 that's why they're part of the consortium,
7 versus a model that says a consortium is
8 building a solution that, because it's
9 against a common set of standards that
10 others outside the consortium might share --
11 might be applicable to others outside the
12 consortium. And so how do you make sure
13 that it is it's available and not being in a
14 consortium doesn't mean you can't use these
15 other solutions. Maybe states never join a
16 consortium, and at the end the look around
17 and say, You know what, that's the thing I
18 think is best -- I want to -- I'm getting on
19 that.

20 MR. NICHOLS: It does bring up --

21 MS. WEISS: And is that okay for
22 us to design it that way? So it's two
23 different models of design more than it was

1 about productizing per se I think.

2 MR. NICHOLS: It does bring up an
3 interesting question into intellectual
4 property. I mean, you're funding these.
5 Why don't you own them then after they're
6 developed.

7 MS. WEISS: No, we will.

8 MR. NICHOLS: And so anybody
9 could use them after that.

10 MS. WEISS: Right.

11 MR. NICHOLS: That would be
12 great.

13 MS. WEISS: Right. Yeah.

14 MR. FISHER: Let me focus just on
15 the idea of developing a common Algebra II
16 test that matches the standards that are
17 being talked about. Now, at the political
18 level that sounds like a great idea. At --
19 when you start following down the rabbit
20 trails you start to run into problems real
21 quick.

22 For example, in order to build a
23 test you have to have test specifications,

1 you have to have items, you have to have
2 test blueprint. But suppose you go through
3 all those steps and you've got this nice,
4 tough-looking test, and then it probably
5 will be tough because those content
6 standards that are floating out there are
7 tough.

8 Somebody's got to set cut-scores.

9 So you go through some process and set
10 maybe achievement levels or, you know, a
11 passing score or whatever you want to call
12 it. Given the current skill level of our
13 high schools who have not been exposed to
14 that kind of instruction, the impact will be
15 horrendous; the headlines will be
16 horrendous.

17 Secondly, somebody's got to
18 collect the information to demonstrate that
19 those cut-scores or proficiency levels have
20 validity related to whether or not the kid's
21 going to succeed, presumably in his first-
22 semester math class at college.

23 We already know that 50 percent

1 of the kids who take placement tests fail.
2 So now we superimpose challenging standards
3 on this challenging content on this current
4 situation, and it's going to produce a lot
5 of very bad press. And in the states that
6 have end-of-course graduation requirements,
7 there's going to be a lot of kids who don't
8 graduate. Bad press, not a good thing to
9 do.

10 If you're going to approach this,
11 I see you've got two choices. One is to set
12 up development of prototypes, show people
13 how to move in the directions in which they
14 need to move and use the power of the bully
15 pulpit and the power of the political
16 leadership to get people to move in that
17 direction.

18 It won't happen overnight. You
19 can't simply drop another set of high
20 standards on people and say, Whoops -- can't
21 do that.

22 MS. WEISS: Right.

23 MR. FISHER: And in any state

1 that has high stakes that's going to try
2 this you could trigger all kinds of
3 litigation. And that's not the thing to do.

4 And there's all kinds of work
5 that's going to have to be done to link
6 together a high school curriculum and the
7 instructional program with what the college
8 expectations are.

9 So, to me, I guess I'm kind of
10 endorsing what Brian is trying to lead you
11 to, if I'm interpreting Brian's comments
12 correctly. We're better off teaching people
13 how to do what we want them to do and show
14 them the prototypes and help them develop
15 and help them learn new skills and show them
16 how we can get there rather than simply say
17 two years from now -- boom -- we're going to
18 have a new common test. And we're going to
19 have all this massive upheaval of negative
20 publicity when people don't do well.

21 MS. WEISS: Okay. Any other
22 thoughts on this question of number of
23 different types of tests and different areas

1 or how consortia might be?

2 MS. SISKIND: This is a thought.

3 It's not an answer. I'm really stating the
4 obvious. I think there's more room for
5 innovation in the other components of the
6 system than this summative common test. I
7 mean, we know how to do that pretty well.
8 It's the development of the other --

9 MS. WEISS: The performance
10 tasks --

11 MS. SISKIND: -- components --
12 yeah.

13 MS. WEISS: -- and those things.
14 So maybe let's take it from here to a
15 discussion of the LEA's role in this grant
16 and where we see -- what would be effective
17 uses of LEAs, teachers, time, funds from the
18 grants in support of some of these different
19 pieces that we've talking about. Where
20 would you focus that effort?

21 MS. SISKIND: Well, I think one
22 thing that would be attractive is in
23 professional development, even from the

1 state perspective. Because when funds are
2 cut that's the kind of thing that's cut. I
3 mean, we're really short on funds for
4 instructional development, meaning the
5 development -- these are really
6 instructional components -- those formative
7 components that we're talking about. So,
8 you know, I think that would be some place
9 that both states and districts would welcome
10 some funding.

11 MS. WHALEN: Can I just ask a
12 clarifying question to that -- so
13 professional development around the
14 formative point. Do you mean in developing
15 them or scoring them, how to use them,
16 ongoing professional development, or kind of
17 one-time or kind of short-term? How do you
18 develop the system capacity to do this well?

19 Or are you trying to kind of backfill cuts
20 in the system?

21 MS. SISKIND: Well, I think when
22 I'm talking about the classroom assessment
23 I'm talking about questioning techniques

1 that you would use in the classroom every
2 day -- that kind of thing. That would be
3 wonderful if we could include in our
4 professional -- our college preparation
5 programs, but we don't. And so that's one
6 component that I think is key.

7 So it's more on the instructional
8 side -- it's instructionally embedded --
9 kind of moment-to-moment assessments. And
10 that is slow going and expensive, and you
11 can maybe do 25 teachers at a time -- a
12 hundred teachers at a time and so forth.

13 The other thing I think is in
14 these more extended kinds of activities --
15 that we've described them in different
16 ways -- but having development in terms of
17 development of those and scoring of those
18 and, you know, that kind of thing --
19 specification of those.

20 MS. WEISS: So, Carol, assuming
21 that there's clearly ongoing costs once a
22 system like this is up and running in the
23 scoring and moderation and teacher

1 involvement, what do you see as the up front
2 investment that we would need to make, maybe
3 even state by state, in putting together the
4 systems that in an ongoing way are used to
5 train teachers, oversee these scoring
6 events, programs, the moderation or auditing
7 vehicles that need to exist? What are the
8 pieces of that that we should think about
9 that would need to be put in place?

10 MS. CAMPBELL: Well, I think it
11 does come back to the teacher involvement in
12 summative assessments and the balance
13 between open and closed responses, as well
14 as the ongoing day-to-day assessment for
15 learning in the classroom.

16 So on the summative pieces, which
17 would be teachers involved in scoring, open
18 responses on end of year tests -- perhaps
19 interim assessments -- I mean, the way that
20 I know it's been done is fairly intensive
21 work, quite often over the summer, which is
22 why I'm saying it's additional work, it's
23 voluntary work, it's paid work -- it's not

1 all teachers in the system, it's something
2 that a group of teachers choose to get
3 involved in.

4 So in Ontario it's over a 1,000
5 teachers a year that volunteer and then are
6 contracted to do this work. They have
7 intensive training over, you know -- short
8 but intensive periods over a few weeks in
9 the summer and they come together, then
10 there's some support through the year. Now,
11 that's particular to the external examiners
12 or in these types of assessments. And then
13 it just keeps going each year as different
14 teachers become involved in each year and
15 some are returning.

16 So, yeah, there's certainly
17 dollars involved. There's human
18 resources -- financial resources. But it's
19 that intensive type of training and
20 development.

21 For the ongoing assessment for
22 learning we've done it through a whole range
23 of ways. So in Ontario we have a government

1 that's absolutely committed to education and
2 teacher capacity building as the main driver
3 of change.

4 So we did the big provincial --
5 pull teachers out and everybody gets trained
6 so-called in literacy and in numeracy and
7 summer programs and release days -- with
8 mixed success in that we all know that, you
9 know, pulling out and intensive training has
10 mixed success.

11 So just as Teri's saying, we're
12 much more into job-embedded professional
13 development, professional learning
14 communities, resources into classrooms, web
15 casts, pod casts, literacy coaches, all of
16 these types of support.

17 So it actually becomes part of
18 the routine of the work in our schools and
19 districts. There is release time, but
20 there's also time built into school meetings
21 and school processes. So it's hard to sort
22 of clearly say this dollar was for this,
23 because it's actually part of the

1 development of the whole school now.

2 MS. WEISS: Somebody asked us
3 whether -- how many teacher work days and
4 professional development days are scheduled
5 each year.

6 MS. CAMPBELL: In Ontario we --
7 this is where nobody's going to be
8 surprised. We -- provincially we have two
9 professional development days, which sounds
10 tiny compared to what I've been talking
11 about. But there two days that we see as a
12 province are dedicated to provincial
13 priorities. But then at the school district
14 level there is at least another ten days
15 that are used. And because they're on very
16 similar priorities, it's those pieces.

17 And so there are release days
18 that are part of the school year. But that
19 kind of hides the fact that there's a whole
20 range of release built into program
21 initiatives, policy initiatives, curriculum
22 initiatives. But increasingly we're not
23 pulling teachers out much any more; we're

1 providing times within schools for team
2 working or for work across families of
3 schools.

4 So networks where teachers across
5 schools will get together every six weeks
6 and share student work, assess the student
7 work at moderating, marking, and levels. So
8 we built in that as part of the local piece
9 rather than something that's at the
10 provincial level.

11 MS. WEISS: Thank you. Tom?

12 MR. FISHER: Just a reminder that
13 our political leaders really like to have
14 quick turnaround of data. And No Child Left
15 Behind requirements as they are now kind of
16 dictate that you've got your score
17 processing done by August, at least, so you
18 can turn in your reports. So in the back of
19 your mind you have to sort of watch what's
20 going on with the No Child Left Behind
21 requirements now that relates to all this
22 other stuff.

23 One of the best things that ever

1 happened from this standpoint when I was in
2 Florida was we created some contracts with
3 universities and the school districts that
4 actually set up test development centers.
5 And teachers were assigned from the school
6 district to center, and they worked there --
7 six months, three months, a year,
8 whatever -- and then they were sent back to
9 the school. When they went back to the
10 school they really had topnotch training in
11 the kind of stuff that we're talking about.

12 So I can envision how part of
13 your requirements here would be that the
14 states would have to set up those kinds of
15 arrangements. Teachers don't like to be
16 doing this kind of stuff at night and on
17 their own time, and it's really not fair to
18 them.

19 MS. WEISS: Sorry. Go ahead,
20 Paul.

21 MR. NICHOLS: There's a nice
22 relationship again between North Carolina
23 State and teachers in North Carolina that

1 has that sort of -- where the teachers and
2 the university both benefit because the
3 teachers are getting training in the latest
4 understanding of teaching and learning,
5 whereas the -- and the researchers are going
6 out there in the classroom and they're
7 gathering evidence and research about how
8 teachers teach and how students learn and
9 what works and what doesn't. And it's
10 mutually beneficially for both groups.

11 MR. GONG: I just started
12 thinking about this in a little more detail,
13 and so let me tell you the way I'm thinking
14 that I've come to what the recommendation
15 is.

16 If there's \$175 million and you
17 divide it to 25 states, just as a point to
18 think about this, that's \$7 million a state,
19 which is 35,000 teachers would each have two
20 days of work funded at \$100 each if you
21 funded their release time. I mean, this
22 is -- you know, so you cannot think of this
23 as direct funding, because I don't think it

1 will be anything. So that's why I think --

2 MS. WEISS: If it's distributed
3 like --

4 MR. GONG: Right.

5 MS. WEISS: -- pixie dust.

6 MR. GONG: Right.

7 MS. WEISS: Right.

8 MR. GONG: Right. So you have to
9 think about what the -- so the first one I
10 think you have to put out some pretty strict
11 guidelines about what this really is used
12 for and ask the states what they -- if they
13 got \$7 million going -- that was going to
14 districts what would actually leverage that.

15 Because I thought of three
16 things -- and I thought Ann's question was a
17 great one. If you're developing system
18 capacity on a one-time basis then actually
19 that's pretty good. If you're doing
20 operational things it's really bad.

21 MS. WEISS: Right.

22 MR. GONG: So you've got to think
23 about what it is you're trying to do. And

1 the other is if you're trying to leverage it
2 are you providing -- are you funding
3 leadership for things that are being funded
4 some other way? Are you putting in seed
5 money that someone's going to have to match
6 in either money or in kind? Are you trying
7 to fill in what wouldn't get done -- or
8 speed up?

9 So I think it's possible to have
10 some ideas about what the Department
11 things -- and you can get advice about
12 this -- I think go to the states and say, If
13 you had \$7 million and what you've heard
14 today what would really make a difference?
15 And I think you'll get some pretty good
16 advice about how to structure that.

17 And it won't be the same for
18 every state. Some states will say, We've
19 got a bunch of P.D. days -- just help us
20 direct it. Other people are going to say,
21 Help us develop our university ties by --
22 we've got a great university system with
23 relationships already there, let's just

1 expand that or something. I think it will
2 vary by state.

3 But I think that just looking at
4 the amount -- it sounds like a lot, but I
5 think it has to be a leveraged type of fund.
6 It can't be different.

7 MS. WHALEN: So can I ask a
8 follow-up question to that? I think --
9 aside from the teachers you touched upon a
10 couple of other kind of interwoven systems
11 and relationships that are all attached to
12 the assessments. We talked about
13 curriculum, professional development,
14 University partnerships, scope and
15 sequence -- all these things.

16 What should we be thinking about
17 when we're writing this notice in terms of
18 state expectations for changed management
19 within their systems? What should be some
20 non-negotiables that state would -- at the
21 consortia should come to the table with to
22 think I'm going to change this to align with
23 that? Or should we just be flexible and

1 ask the open ended question in order to make
2 this work to align with your theory of
3 action what will you change? But what is
4 the right way to frame this and what in your
5 vision is a non-negotiable for this?

6 MS. WURTZEL: Go ahead, Brian.

7 MR. GONG: I think that if you
8 have something like Carol says -- we've got
9 an idea about what it takes, and this is
10 sort of like the assurances that were put
11 out for the Race, but those were not
12 capacity things; those were more doing
13 different things.

14 You said -- we think that in
15 order for this to be successful you're going
16 to need these types of things. Can you show
17 us what you have there and how you're going
18 to spend your money. And when a state does
19 that -- or a set of states -- they'll be
20 putting in \$20 to every one, because the
21 things that are going on here are not being
22 funded by assessment money or even federal
23 dollars.

1 And I think this goes along with
2 my general philosophy. States are very
3 willing to commit their money to things that
4 are worthwhile doing. You don't want to be
5 in the position of saying, I'm only doing
6 this because of the Race to the Top money,
7 and if the money goes away, I'm stopping. I
8 think it's -- I think you really want to
9 avoid that type of thing.

10 So states that are invested --
11 states will invest their dollars in doing
12 things that they think are really important.

13 And I would ask -- be pretty tight about
14 those.

15 And one model is Massachusetts,
16 and some states are doing this. On this
17 turnaround schools they say, We have a model
18 that we think effective schools have to have
19 these elements. And so when we come in,
20 this is what you have. You have to attend
21 to all of these. If you can show us that
22 you already have something along those
23 lines, you know, you don't have to do ours;

1 you can do yours. But you have to have
2 something and it has to be coherent.

3 And I think that if you have that
4 set of conditions in it and its -- we know
5 that you have to have broad-scale teacher
6 competence and buy-in and things like that
7 then I think that -- and most of the states
8 have it already. This is -- Tom has been
9 reminding us states have been fighting this
10 battle about assessment for over 20 years
11 now. They've developed a lot of credibility
12 with their systems.

13 So I think you're just building
14 on top of that. I don't think you're trying
15 to -- if -- it's one thing to say, you know,
16 state standards -- content standards are
17 really bad and their high school assessments
18 are not credible and so we're trying to help
19 change that. That -- I think you're going
20 to have a lot of push back if you say that.

21 Another way to do it is to say, We're
22 building on what's there and then make sure
23 that they're actually building on it. I

1 think most states are willing to do that.

2 MS. SISKIND: Let me ask Brian a
3 question -- it's kind of a pragmatic
4 question. I don't disagree that the states
5 could tell you what they could use those
6 monies for. But when you join a
7 consortium -- say there's ten states in a
8 consortium -- and those monies are going to
9 be allocated to the districts -- the local
10 education agencies -- how would you handle
11 that in both the call for proposals and the
12 proposals, those differing needs?

13 MR. GONG: I think you're raising
14 a great question. I think the LEA stuff --
15 there will be some things that are common
16 depending on what the -- if someone is doing
17 an innovative thing that few other states
18 are doing, then there will be some common
19 needs.

20 But if for some wonderful thing
21 North Carolina and South Carolina decided to
22 work together, I think your LEA needs are
23 going to be somewhat different, so I would

1 leave that flexible within the grant. I
2 would ask the consortium what are you doing
3 together and then on the local things what
4 are you probably doing separately? I don't
5 know if that --

6 MS. SISKIND: Just allocated
7 proportionately or equally among the
8 partners and then let the partners make
9 those determinations.

10 MR. GONG: Well, I wouldn't have
11 the U.S. Department allocate it. I'd ask
12 for the proposal. Have the states -- so
13 that's why this getting -- letting the
14 states know what you want up front is really
15 important. If you give states a 90-day to-
16 do list that's really, really hard for them
17 to work through all of these details.

18 I wanted to just raise an idea.
19 There are lots of different models for doing
20 awards. There's sort of the award where you
21 have to have the proposal up front and
22 then -- and that's fully worked out and then
23 funded and that's what you're expected to

1 do.

2 The second one is we have some
3 checkpoints, and, you know, that's a pretty
4 different one where this is -- then you're
5 supposed to deliver the next plan or work
6 through it.

7 And then there's a third one that
8 foundations use, which is a much more
9 interactive one, which is where -- we know
10 that some of these things are really hard
11 and no one could have thought ahead of time
12 how this step works out, either because it's
13 too far in advance or it's dependent on how
14 this first part worked out -- so how this
15 consortium actually is working out. And so
16 we're going to negotiate that part later.

17 And I think there's a lot of
18 advantage if you can work that out. For
19 some of these -- if you're working on more
20 problematic or complex ones, the third model
21 would actually be really helpful if you
22 can -- it's very -- it's a little different
23 than a memorandum of understanding that the

1 Department's used before, but that might be
2 useful for some of this.

3 And I think it will help make it
4 so you don't have to solve all those
5 questions up front. If you -- I think if a
6 state said I have to answer all of this
7 before I sign on it's going to take an act
8 of faith before they, you know, are going to
9 say our two states are going to work
10 together. That's really hard.

11 MS. WEISS: So one of the states
12 sent us a question that's right along these
13 lines that said, Would the Department
14 consider funding one or two additional work
15 groups comprised of states, vendor reps, and
16 assessment experts with a subsequent set of
17 pilot studies to determine the best fit for
18 assessment so that basically over time you
19 would slowly build out and adapt an
20 assessment system with transition funding to
21 help the states get there from here.

22 Now, putting aside all the
23 statutory requirements that we're trying to

1 operate under for a minute -- just assume
2 those away -- talk to me about -- that
3 approach is a little different from the
4 foundation one that you were talking about.

5 But do you have sort of thoughts, pros and
6 cons? Because we can at least go back and
7 noodle on them within the bounds of what we
8 can and can't do here.

9 MR. GONG: I don't see how it was
10 different from what you've talked about as
11 long as a state submits the proposal and
12 have some other states agree to it. It just
13 sounds -- I mean, it sounds -- I thought it
14 fell well within what was -- what the
15 statutory requirements were.

16 MS. WEISS: Yeah. Our big
17 statutory requirement is that we have to
18 obligate all the funds by September. And so
19 we have to figure out a vehicle then for how
20 to change courses effectively based on what
21 we're learning mid-stream, which we're sort
22 of working on what the vehicle might be for
23 doing that. It's not the way this is

1 normally done, but I think we all believe
2 that in this case it's probably -- it's
3 certainly the right way to do it and we need
4 to see how close we can get to that.

5 MR. FISHER: Perhaps -- you know,
6 you've got a real time constraint. It's a
7 bureaucratic, legal, fiscal time constraint.

8 And perhaps that reinforces the idea of now
9 biting off more than you chew, you know.
10 Divide the task into smaller units it can be
11 accomplished without agonizing over whole
12 scale upheaval of American education, you
13 know, with such small lead time.

14 MS. WEISS: Now, the funds can be
15 spent down over many --

16 MR. FISHER: Well, I
17 understand -- I understand --

18 MS. WEISS: -- you know, four to
19 five years.

20 MR. FISHER: I understand that.
21 But planning is very important. That's --
22 every speaker here has said you've got to be
23 careful. You've got to think about those --

1 every person said unintended consequences.
2 So it takes time to sit down and think about
3 that -- what are going to be the ripple
4 effects -- I'm not going to repeat my speech
5 or anybody else's speech.

6 But there is some advantage in
7 thinking modestly about it because we're in
8 tough economic times. And if we envision in
9 our collective minds this avant garde, truly
10 advanced, computer driven whoop-de-do
11 assessment and we develop it and turn it
12 over to the states and they, Okay, but I've
13 just cut my state budget 20 percent and I'm
14 laying off people like crazy, it's going to
15 fall with a big thud.

16 MS. WEISS: Carol?

17 MS. CAMPBELL: I'm thinking
18 connected to the question you've just asked,
19 Joanne, and actually, Ann, when you asked
20 your question, one of the things that keeps
21 going through my mind is the governance
22 piece. And I know that this is a bigger
23 piece, but it does slightly play into

1 whether it's too nationally arranged or five
2 consortia or whatever.

3 And I just want to be really
4 clear -- because I'm not sure it came
5 through when people were asking me about the
6 teacher dimension -- teacher dimension is
7 absolutely fundamental, but I don't want to
8 leave the impression that it's our teachers
9 who are our psychometricians and all of
10 these sorts of things.

11 We have an education quality and
12 accountability office to the tune of \$35
13 million a year that has full-time people
14 doing this work with the support of teachers
15 throughout. But, you know, however it's
16 arranged through all the various types of
17 organizations in the U.S. I do think that
18 you need to really work through a
19 sustainable governance piece. You know,
20 similarly in England there's multiple
21 national governance arrangements through the
22 Office for Qualifications or Regulator, et
23 cetera, et cetera. So I just want to be

1 clear that teachers are critical, but there
2 has to be a whole set of supportive
3 infrastructure around it as well.

4 MS. WEISS: Right.

5 MR. NICHOLS: So when you're
6 talking about things that are non-negotiable
7 you've always laid out a vision for us about
8 what you wanted to see, which would probably
9 be the same things you'd want to see from
10 these consortia.

11 I think one of the things you
12 could add to that is something that's
13 achievable and sustainable. They have to
14 argue that what they're doing in this
15 consortium is achievable and sustainable
16 within the time frame that they have. They
17 have to make a strong argument for that. I
18 think that would be something that Tom was
19 talking about in chunks and sort of chunks
20 you can chew.

21 MS. WEISS: So let's talk for the
22 last few minutes that we have about advice
23 that you have for managing consortia,

1 governance within a consortia.

2 So if you're a state trying to
3 put together a consortium, what are the
4 things that you think they need to think
5 about in order for this project to go well
6 to adapt to the things that they're learning
7 over time, and so on.

8 And we'll be listening for the
9 ears of what are the kinds of questions we
10 should ask states, but really this is almost
11 more at this point for the benefit of
12 states -- some free advice on things they
13 should think about as they're forming these
14 consortia.

15 MR. NICHOLS: Almost like a
16 marriage, isn't it, to be able to
17 communicate and get along. So I think there
18 should be states that have common interests
19 in some areas, especially in the area in
20 which their consortium is working towards.
21 Now, if it's ELL, the states can have a
22 common interest or a common approach to
23 ELL -- something like that.

1 MS. WEISS: How about governance
2 issues, leadership management -- you know,
3 anything from the nuts and bolts to the more
4 strategic?

5 MR. FABRIZIO: I would think
6 you'd want to possibly talk with the folks
7 from Achieve about what was learned through
8 the ADP Algebra II Project, because I know
9 for many of us it involved, you know, having
10 to get the Attorney General's Office to sign
11 these memorandum of agreements because the
12 money was flowing through Ohio.

13 And so really the contract in
14 North Carolina was with the state of Ohio,
15 and then that state of Ohio was paying
16 Pearson. I mean, there were all of those
17 kinds of legal things that you probably want
18 to find out what kinds of -- those kinds of
19 things that you need to build into this.

20 And then while I'm at it -- and I
21 don't know if you could have any control
22 over this or not, but I could just see --
23 and I'll say I'm from South Carolina -- the

1 General Assembly saying, Oh, Teri got \$8
2 million to be part of this consortia, so we
3 can cut the testing budget by 6 million and
4 it won't really hurt. So --

5 MS. WEISS: They've already done
6 that.

7 MR. FABRIZIO: So I don't know if
8 there's any way to sort of protect the
9 states that do this kind of work -- that it
10 becomes very clear that it's not an
11 invitation to legislators to cut back on
12 education funding.

13 And, again -- and one other quick
14 thing I did want to throw on the table
15 because nobody's brought it up -- and it has
16 to do with the academically gifted. It gets
17 at that whole issue of grade level
18 assessments.

19 You know, if we're really trying
20 to push kids to their maximum then we're
21 going to have the problem we currently have,
22 which is kids in seventh grade or eighth
23 grade taking Algebra I and they're having to

1 now take an eighth grade math test plus
2 take, at least in our state, the Algebra I
3 end of course test.

4 And it would seem like there
5 should be some allowance for students that
6 have progressed to a higher level to be able
7 to take the higher level test and that would
8 count as their lower level participation, so
9 to speak, so that we truly are, you know,
10 rewarding the kids and not saying, Oh,
11 you're so smart that you have to now take
12 two tests.

13 And I know that the ruling I
14 think was that there were different
15 standards -- different standards for eighth
16 grade math versus different standards for
17 Algebra I. But if one can be viewed as a
18 higher level assessment then that should
19 count. And I don't know if that's something
20 to keep in mind.

21 MR. GONG: This may be repeating
22 some. I think that if the RFP is clear
23 about whether the goal -- whether a

1 requirement of the consortium is that -- the
2 intent is that the states adopt it then you
3 get certain groupings. If it's to develop
4 something then you may have some -- if it's
5 to use and develop them then that might be
6 the overall best thing.

7 I'd encourage states to think --
8 look at the types of multi-state
9 partnerships that have happened in the
10 past -- and some of have been more
11 successful than others. I'll just name some
12 of them, and you might think of others.
13 There are -- many states have had EAG or
14 GSEG experience. You can think about, you
15 know, what went well and what didn't go
16 well. You might say, Boy, I don't think I'd
17 join one that was led by that state again
18 or -- you know.

19 There's the Achieve and Pearson
20 Algebra II one that I think about 18 or more
21 states have some experience in that.
22 There's the new standards project; there
23 were the SCAS projects. NECAP is -- Gaby

1 Dorchek's here from New Hampshire; he can
2 talk to you about that model.

3 Just a little comment about them:

4 I think one of the important parts on this
5 is what is the role that the state plays.
6 So I would say NECAP -- the states provide
7 the leadership and they tell people what to
8 do. And my center's involved with that.
9 We're providing support, but we are not
10 telling the states what to do. And they
11 hire a contractor and they tell the
12 contractor what to do. And the states own
13 the stuff and they are -- they control it
14 over a long period of time.

15 I would contrast that with the
16 Achieve Pearson one because actually Pearson
17 owns the test. And so states had some
18 involvement at the beginning. Achieve had a
19 key role in convening the states and
20 providing the intellectual leadership.

21 But if you come into the Achieve
22 project now you wouldn't know about this
23 multi-state involvement in the test

1 specifications and stuff because when you
2 join now you don't have any say in that --
3 you're buying a product. Although someone
4 mentioned this validity studies and others,
5 so there are a lot of very nice supporting
6 things that Achieve is still driving. And
7 the states have a counsel and things like
8 that.

9 So that's a little different
10 model. And I think it's important about
11 whether you want to end up with -- whether
12 you want to own the thing or whether you
13 say, you know, I would be glad to let
14 someone else sort of take care of the
15 headache.

16 The new standards project is
17 different from Achieve Pearson, because new
18 standards didn't have a contractor, you
19 know. So they tried to have the expertise
20 within their group. And if you think about
21 developing a common assessment then the
22 states are all going to -- or the states are
23 going to work with not -- new standards was

1 not -- well, states were working but the NCE
2 was responsible.

3 So that's a third party that's
4 providing all the technical leadership. And
5 so think about what third parties that are
6 vendors or not vendors might be able to do
7 that. People know about the SCATS Project.

8 So -- and you're working with
9 vendors now -- think about the vendors'
10 strengths. And if you've gone through more
11 than one vendor you probably didn't do it
12 because you thought they had a lot -- the
13 first one had a lot of strengths. But I
14 don't know -- maybe it was just a purely
15 budget thing.

16 But part of the problem is I
17 don't know many states who had experience
18 with multiple vendors in a positive way. So
19 it's hard for states to know what the
20 various strengths are that different vendors
21 bring. You may have to have to -- you may
22 want to talk to other states about that type
23 of thing.

1 So, anyway, there's a -- I think
2 that states -- if you're going to start
3 working together start thinking about what
4 it is that you want to do, who you want to
5 work with, both states and non-states -- and
6 starting now is not too early.

7 MR. NICHOLS: You may have
8 already -- this may already be included in
9 your thinking, but an evaluation team with
10 an action theory and --

11 MS. WEISS: Uh-huh.

12 MR. NICHOLS: It's already in
13 there?

14 MS. WEISS: Uh-huh.

15 MR. NICHOLS: Just checking.

16 MR. GONG: Let me just add this
17 to the team. So we talked about all this
18 process and structure. If I were in a team
19 I would say do I agree with them
20 philosophically? Do I trust them? Are they
21 competent to deliver what they're going to
22 deliver? And are they -- who's providing
23 the intellectual leadership on this?

1 So if I'm the state -- and I
2 think I am -- then I'm looking for other
3 people. If I'm not providing that I'd
4 better find someone who is going to provide
5 it and that I really trust. That might be a
6 contractor, it might be another state, it
7 might be a group of states.

8 But -- and then sensitivity and
9 shared values about these operations things
10 because in making a group effort work -- and
11 you all knows this -- it takes a lot to make
12 this a successful thing. And this is a
13 really important project. I mean, it will
14 be important to you for a lot of ways.

15 I'll just mention something I
16 know about the NECAP states. The NECAP
17 states meet -- the assessment directors meet
18 every week. They meet together about --
19 okay -- every two weeks -- sometimes phone
20 call, some in person. That is not a model
21 that most states can do. These are four
22 states that are close together
23 geographically. But the reason that they do

1 that is because they have to iron out a
2 bunch of decision about things and it takes
3 that type of communication.

4 Now, if you're buying a turnkey
5 system -- if you're saying, Just give me the
6 thing and I'll adopt it and you tell me what
7 to do with it, that's one thing. If you
8 are -- feel like you have to have control
9 you'd better have a different model than a
10 turnkey one about adopting things.

11 So, I mean, look at your own
12 state and figure out what's really going on
13 in there. And talk to the folks from Maine.

14 They came a little bit different -- they'll
15 tell you about how painful it is in some
16 ways and what the advantages are.

17 But I would really encourage you
18 to talk to people -- states that have had
19 these experience and draw on your own
20 because it's not just what you propose, but
21 I think it's really key about who you're
22 going to work together and make sure that's
23 really successful.

1 MS. WEISS: So let me just ask a
2 follow-up question because you sort of --
3 you definitely started addressing this,
4 Brian. But a consortium doesn't have to be
5 made up of members who are all co-equals
6 with all the same responsibilities. You can
7 have a lot of differentiation.

8 What are the different types of
9 differentiation of roles that might -- or
10 differentiation of participation that might
11 be worth thinking about? You talked about
12 it a little bit when you said, you know, you
13 might decide that you're going to go with
14 somebody who was providing the intellectual
15 leadership. What are the ways you might
16 sign up and the benefits to doing that?

17 MR. NICHOLS: When you talk about
18 intellectual leadership, you know, I really
19 think you -- you really need to look to your
20 university partners. NSF, Department of Ed
21 have funded a lot of research over the
22 years. They've built up a lot of expertise
23 and experience on assessment and curriculum

1 and how they work together.

2 So there are some groups out
3 there and labs and senders that have some
4 great expertise. And I think I would
5 encourage any consortium to include
6 university partners.

7 MS. WEISS: So let me get back to
8 my differentiation question. In other
9 words, does it make sense to have sort of a
10 core group of one or two or three states
11 that are providing the leadership and other
12 states who come on and say, you know what,
13 I'm giving you my proxy vote for this stuff
14 and here's the way in which I want to be
15 involved -- and it's sort of a broad
16 consortium almost of people who are
17 following and a small group of people who
18 are leading. What are models that you think
19 might or might not work there?

20 MR. FABRIZIO: It would seem like
21 that's something the different states could
22 work out among themselves as far as the
23 proposal that they'd put together. I mean,

1 if one of the states says, I want to
2 participate, but here's all I'll be able to
3 do, then if that's okay with everybody else
4 in that consortium that seems like that
5 should be their prerogative.

6 MR. NICHOLS: Maybe all you'd
7 have to do is to make them have clear
8 decision-making rules in their proposal and
9 then they could decide whatever they wanted.

10 MR. GONG: I clearly haven't
11 thought enough about this question. I
12 thought about a related one that I'm trying
13 to put on the U.S. Department of Ed. So
14 that's really not answering your question.

15 I'm great friends with folks from
16 Massachusetts. My question is how can I --
17 how could I get Massachusetts to work with
18 any other state? And so -- and what it is
19 is like -- so the U.S. Department of Ed, if
20 you want states to work together and you
21 care about what the quality is then I think
22 that if you let states self-organize any
23 number of things could happen. But if there

1 are certain things that you want to happen I
2 think that it's helpful if you help make
3 that happen.

4 And so here are two models. One
5 is, if I wanted -- I want to join with Lou
6 because I know is really smart and I know
7 that he's really to work with and he has
8 really --

9 MR. FABRIZIO: Okay. How much do
10 I owe you?

11 MR. GONG: Good track record.
12 But Lou may look at me and say, You're not
13 so smart, you don't have such a great track
14 record, and -- you know. So -- and states
15 may be looking around and they say, It's
16 like will you work with me and we have these
17 problems of everyone wants to work with
18 someone better than they are and no one
19 wants to work with someone weaker than they
20 are.

21 I don't think that that's a good
22 model -- I don't think that -- if you let
23 that happen the consortia will not work in

1 the long run. So I think you have to think
2 about what a good mix would be and try to
3 encourage -- I don't know if you can require
4 it, but I think there are some other things
5 that try to -- if I were a state I would
6 first of all figure out what it is I wanted
7 to do and then I'd make sure that I had
8 people that I agreed with and people that I
9 didn't -- that didn't agree with me but that
10 we could come to agreement and within the
11 project time, because if I work together
12 with only people who are like what I am, I
13 don't think I'm going to get a good
14 product -- or as good a product as if I were
15 challenged in some way.

16 An example was the NECAP states
17 had a time with the science assessment. And
18 the three science directors all had very
19 difficult views of what inquiry was. And I
20 think that by talking together they came out
21 with a better view of what science inquiry,
22 and the assessment was better.

23 I think that's actually a really

1 good example of how multi-state working
2 together was better than what any one state
3 could do. But I think that was -- it was
4 good that that happened that way.

5 MS. WEISS: So are you offering
6 to be the yenta matchmaker of consortia?
7 I'm just trying to understand the --

8 MR. GONG: No, I'm not offering.
9 But I think it almost takes some personal
10 brokering to make it happen.

11 MR. FABRIZIO: Well, Brian, I
12 hate to burst your bubble but I think that
13 brokering is probably already starting to
14 occur at the state commissioner and
15 superintendent level. And I think they're
16 doing it with basketballs and, you know,
17 whoever gets it in -- they become the
18 captains and then they get to pick.

19 MR. GONG: There's a lot of
20 inquiries going on about teams. Vendors are
21 doing it already. Commissioners are doing
22 it. Some state assessments folks are doing
23 it. You can imagine the strengths and

1 weaknesses of each one. I think -- I'm just
2 pleading saying if we know that there are
3 better ways to do it don't leave it up to
4 the sort of natural politicking that's going
5 on.

6 MS. WEISS: Somewhat delicately
7 put. Thank you. On that note we're going
8 to take a 15-minute break. Before we do I
9 want to just say a couple of quick things.
10 Anybody who is going to be a public speaker
11 in our next session please see Anya. She's
12 going to tell you what to do. So instead of
13 taking a break go see Anya if you're
14 speaking.

15 The rest of us will convene back
16 in this room in 15 minutes at 3:45. And let
17 me also just really give a round of applause
18 to our experts and thank them so much for
19 sharing their wisdom with us. So thank you.

20 We'll see you back here in 15.

21 (Whereupon, a short recess was
22 taken.)

23 MS. WEISS: I'll ask you to take

1 your seats. And let me just share with you
2 how this will work. There are a number of
3 people who signed up ahead of time to be
4 speakers at different events that we're
5 hosting around the country.

6 And we have five people who have
7 signed up to be speakers today. They'll
8 come up in order and they'll each have five
9 minutes to speak. And for those of you who
10 are speaking on your podium you'll see a
11 timer so that you'll know how you're doing
12 and how long you've taken and how much time
13 you have left. And the handy-dandy
14 lights -- you might have noticed them
15 working up here -- will turn to yellow when
16 you have two minutes left and red when
17 you're out of time.

18 So, with that, let's get started.

19 And whoever is the first speaker, come on
20 up.

21 MR. TRUBY: Thank you. It's hard
22 for a recovering politician to say anything
23 in five minutes. I'm Roy Truby. I'm senior

1 vice president for Renaissance Learning.

2 And I would first disclose that
3 we do have an interest at Renaissance
4 Learning in that we have interim assessments
5 in Star reading and Star math and Star early
6 literacy in over 44,000 schools.

7 And we are presently working to
8 build an assessment system called the Star
9 Constellation, which will be aligned to the
10 common core state standards. We're doing
11 pretty much what the chiefs and the
12 governors are doing. We're looking at the
13 Achieve and NAEP frameworks and college
14 placement frameworks and others.

15 My current position, though, is
16 really not what I'm going to be speaking
17 from. My testimony is based on the fact
18 that I -- or on my experience as state
19 superintendent in two states and executive
20 director for the National Assessment
21 Governing Board for 13 years.

22 To stay within the five minutes
23 I'm just going to touch lightly on a

1 balanced assessment system development
2 considerations and the balance really
3 between multiple choice and constructed
4 response.

5 You've said a lot about a
6 balanced assessment system so I don't need
7 to say a lot more. I think though that all
8 of us realize that we've had about eight
9 years of giving a test at the end of the
10 year and then trying to figure out who to
11 blame.

12 And we need a balanced system
13 that goes beyond the summative test that
14 includes interim and formative tests that
15 will help teachers accurately assess
16 learning and use assessment information to
17 adjust instruction and improve achievement.

18 For states to provide a balanced
19 assessment system though I think Lou had it
20 right when he said it's going to take more
21 leadership than regulation. And when you
22 think of the half the money that's going to
23 flow through it's really going to have to be

1 part of a rational system, not just sending
2 the money to the districts and letting them
3 do their own thing.

4 There's also I think the issue --
5 well, just follow the logic here. As we
6 develop a balanced system we're going to
7 have more interim benchmarking
8 assessments -- and those are given five
9 times a year to once a month.
10 And the formative assessments are given
11 almost daily

12 So we're going to have more
13 testing, not less, and right now we have
14 teachers who feel like they're drowning in a
15 sea of mandated tests. So how much sense
16 does that make and what kind of feedback are
17 we going to get?

18 I think the only realistic
19 response is that we're going to have to have
20 computer adaptive technology for the most
21 part and get rid of the paper and pencil
22 tests. We're going to have to have shorter
23 development time, a really good balance

1 between multiple choice and machine scored
2 constructed response whenever possible,
3 efficiently administered tests lasting no
4 more than 15 or 20 minutes, and the capacity
5 to generate results for teachers almost
6 immediately.

7 I like Brian's approach too when
8 he talked about thinking about the things
9 that we know how to do and that we can get
10 done in the next two or three years and the
11 things that are really going to take a lot
12 more time -- maybe five to ten years.

13 That may be a way to think about
14 parceling out these consortiums. And no
15 disrespect to universities, but I'd put
16 those into five to ten years block rather
17 than things that need to be done in three
18 years. I think governors and Congress and
19 many folks are really going to be impatient
20 to see results much sooner.

21 So my time's almost running out,
22 so I'm going to think about development
23 considerations just for a minute. When I

1 was executive director for the board we were
2 given exclusive authority over the voluntary
3 national test. I never before or after had
4 exclusive authority over anything. And \$10
5 million later and two years we were just off
6 the ground.

7 So, again, it's going to be very
8 important that we think in terms of the long
9 term and the short term and how we divide
10 that work up. And I won't get to the last
11 point. I've got eleven seconds left. So
12 thank you very much though. This is an
13 important time.

14 MS. WEISS: Thanks.

15 DR. LAZER: Good afternoon. I'm
16 Steve Lazer, vice president, research and
17 development at ETS. My colleagues at ETS
18 asked me to say a few things. I'm going to
19 skip the comments and get -- the charming
20 comments and get straight -- because they
21 never work anyway -- and get straight to the
22 recommendations for R&D stuff.

23 One, I personally like the idea

1 of intermittent summative assessments as
2 part of the summative system. I think it's
3 likely that the system will still require
4 end of year testing events, especially at
5 grades three through eight and at the end of
6 high school courses.

7 However, these don't have to be
8 the only elements of the system. And it's
9 possible to use data collected over the
10 course of the year through one of several
11 models.

12 A simple one, sort of NAEP strung
13 out over time, is take a long test,
14 including performance tests, and break it
15 into piece and give those over the course of
16 the year.

17 A second variant is a sort of
18 more general final with summative tests that
19 cover part of the content given over the
20 course of the year, perhaps weighted less
21 than the final. And lest anybody think this
22 radical every school in the world does this.

23 A different approach that hurts

1 comparability a little more but is still
2 fascinating is to have students complete
3 standardized projects over the course of the
4 year. If development and scoring is
5 carefully controlled I think that you can
6 use these as part of a summative system.

7 Now, of course -- and here I will
8 answer a question a little more aggressively
9 than some of the panelists -- any of these
10 models to me assumes some degree of control
11 or connection over the scope and sequence of
12 instruction. I don't think you could have
13 things happening in the middle of the year
14 if you don't know what's happened first.
15 Whether that's politically or operationally
16 tenable I leave to you folks.

17 Two, use the range of exercise
18 types needed to measure emerging constructs.

19 We always start by talking about items --
20 multiple choice, open ended. I think we
21 should figure out what we want to measure
22 first. But it's quite clear to me from the
23 discussions that we are going to have to

1 move beyond straight-forward, simple
2 multiple choice.

3 I think exercises are certainly
4 going to include, but not be limited to,
5 scenario based tasks, long and short
6 constructed response, tasks that involve the
7 exercise of technology skills as part of
8 constructs and not to measure other
9 constructs, and simulations.

10 Now, one interesting set of
11 questions that got discussed a lot is how to
12 score these. So I'm going to gloss my
13 comments on this a bit. I think we should
14 try to electronically score them but not
15 overlook the professional development
16 opportunities afforded by teacher scoring.

17 But if we say that we have to
18 make fairly sure that scoring is
19 professional development and not drudgery.
20 I'm not sure how many people in this room
21 have been involved in large-scale essay
22 scorings but it's easy to slide. The first
23 50 papers are professional development if

1 we're not careful. The next 8 million may
2 not be.

3 One closing note about innovative
4 item types -- and here from Paul Nichols'
5 perspective this is my Empire Strikes Back
6 comment. We should not repeat mistakes we
7 made in the eighties and nineties. We can't
8 create a system so expensive as to be
9 operationally untenable, and we can't act
10 like the technical challenges associated
11 with performance tasks in particular are
12 mere inconveniences.

13 There are high stakes attached to
14 some of this, and validity, fairness, and
15 reliability really do matter if we want to
16 make decisions about students, teachers,
17 schools, and school systems that are
18 meaningful, defensible, and equitable.

19 Now, this is not to say that
20 technical concern should always win out over
21 all others -- that you must do something
22 that's higher than .9 reliability, but
23 rather we should carefully consider the

1 implications of various decisions so we're
2 not caught by surprised and so we can build
3 research activities around addressing these
4 issues.

5 Finally, if we're going to use
6 these items let's make sure they live up to
7 their information promise and not get kids
8 to do something for a half-hour and then
9 give them like a 3 or something like that.

10 Last, I think this should be a
11 computer-delivered system. I think the
12 emerging skills can only be measured on
13 computer. A lot of people will tell you
14 states don't have the infrastructure yet. I
15 would rather wait than get it wrong, and I
16 think parallel systems mean you won't
17 measure the right skills and content.

18 Again, glossing forward, I also
19 like adaptivity because I think we're
20 talking about harder standards, more
21 rigorous standards, yet, in the end we want
22 a system that tells us about what all
23 students know and can do, including those

1 who may in the foreseeable future not reach
2 these standards. Adaptivity will let us
3 help do that.

4 I think adaptivity will work
5 better frankly if we allow adaptivity to
6 select a bit of off-grade content. NCLB
7 hasn't been friendly to that. I think for
8 both, by the way, lower and higher
9 performing students. NCLB hasn't been
10 friendly to that. I think we should fix
11 that.

12 One tension I think is resolvable
13 is adaptivity versus items that require
14 human scoring. And I don't think these
15 standards will be measured only by machine
16 scorable items. I think we can solve that
17 problem.

18 Adaptation does not have to
19 happen item by item. You could have stage
20 adaptive systems where performance on
21 machine scorable components is used to
22 select open ended exercises that make sense
23 for that student and maybe will even get

1 better bang for the bucks out of the open
2 ended items. Right now, frankly, we assign
3 a lot of things that may or may not be
4 approachable for many of the students.

5 In closing, let me thank you for
6 a real neat event. Let me say we have an
7 unparalleled moment of opportunity here.
8 We've got a lot of advances and the
9 Government's willing to invest in them.
10 However, we also shouldn't blow it because,
11 at least in my career, we probably won't get
12 another chance. Thank you.

13 MS. WEISS: Thanks so much.

14 MR. HUBBARD: Thank you. My name
15 is Jeff Hubbard. I'm the president of the
16 Georgia Association of Educators. I'd like
17 to speak to you today about what GAE
18 believes in in regards to assessment.

19 First, we believe that the
20 definition of formative assessment must
21 include feedback not only to teachers but
22 also directly to students in order to help
23 them understand their own areas of

1 accomplishment and of learning needs.
2 Teachers must be involved in the process of
3 developing high quality assessments from the
4 beginning, as was stated this morning by Dr.
5 Fabrizio.

6 Formative assessments should be
7 designed to provide in depth information for
8 teachers and individual students in order to
9 guide future instruction and learning. We
10 do remain concerned about the usage of
11 student assessment data and how it will be
12 linked to student achievement and growth
13 data to teachers and principals or the use
14 of student achievement as a significant
15 factor for teacher evaluation. However, we
16 do believe that it should be part of an
17 evaluation system, but not the end all.

18 In contrast, it's our belief that
19 interim assessments, while they can provide
20 aggregated data for teachers, still have
21 limited value for the purpose of improving
22 future learning amongst individual students.

23 In fact, in many local systems they're more

1 summative in nature and unfortunately
2 usually amount to an alternative version of
3 a state accountability assessment. It is
4 our belief that the Department should focus
5 on developing high quality formative
6 assessments and reduce the emphasis,
7 therefore, on interim assessments.

8 We further believe that the
9 components of an effective and complete
10 assessment system include formative and
11 summative assessments as well as
12 professional development for educators on
13 how to use assessments to improve
14 instruction and student learning.

15 We also believe in the need for
16 professional development, as stated by Drs.
17 Fabrizio, Campbell, and Gong today, that
18 centers on teachers' examination of
19 students' work from a formative assessment
20 standpoint. This type of professional
21 development is an essential component of a
22 complete assessment system.

23 We also feel that allowing them

1 this professional development will make them
2 look closer and more -- and further into
3 their work to see how they are best reaching
4 their students, and if they are not reaching
5 their students for us to find ways to
6 remediate their practice so that student
7 achievement can be raised along with teacher
8 effectiveness.

9 We also support state and
10 possibly a nationally longitudinal data
11 system to link all 50 states across state
12 agencies. We believe that data should be
13 accessible, analyzed, and used to
14 communicate data to all stakeholders to
15 promote continuous improvement. I should be
16 able at any point to go on here in Atlanta,
17 Georgia, if I have a student transfer from
18 Albuquerque, New Mexico, and see where they
19 are at.

20 Finally, we would hope that the
21 Department would also support building the
22 capacity of all stakeholders to use that
23 longitudinal data for effective decision

1 making. Our parent affiliate, the National
2 Education Association, has put out a
3 conceptual accountability framework of
4 eleven bullets. In the interest of time I'm
5 not going to read those eleven to you. It
6 will be in my written comments.

7 But it does three primary things.

8 It assesses and promotes student learning
9 and identifies those quality schools and,
10 most importantly, it closes the achievement
11 gap.

12 The last thing I'd like to say is
13 actually a concern. It's regarding
14 criterion A(1)(2)(C). In the state of
15 Georgia right now your education stakeholder
16 organizations are being shut out in regards
17 to the task force that is currently being
18 considered.

19 So it would be my hope that as
20 other states that are represented and as
21 this process moves forth that the
22 educational stakeholder organizations which
23 represent over 135,000 educators in the

1 state of Georgia would be allowed to have a
2 seat at the table because when these
3 assessments are done -- when Race to the Top
4 is finished -- whatever states get it -- we,
5 the practitioners, are going to be the ones
6 implementing the program and ultimately are
7 the ones who will be held accountable.

8 So we're asking, whether it's
9 Georgia, Oklahoma, Ohio, New Hampshire --
10 please let us in because we're the ones who
11 have the techniques, we're the ones who have
12 the experience, and we're the ones who are
13 going to be reaching out to the children.
14 So please hear our voices.

15 With that, I thank you for your
16 time. And thank you for allowing me to be a
17 part of this discussion today.

18 MS. WEISS: Thanks so much.
19 Next?

20 MR. WEEMS: Ditto.

21 MS. WEISS: You have five more
22 minutes though.

23 MR. WEEMS: Good afternoon. My

1 name's Joe Nathan Weems, Jr. I am a
2 kindergarten teacher here in Atlanta public
3 schools at Parkside Elementary School. And
4 I first want to thank you for allowing me to
5 be a fly on the wall, but now I get the
6 opportunity to buzz in your ear a little
7 bit.

8 Yes, we are a 21st century school
9 at Parkside Elementary. This is the 21st
10 century, and I think we need to stop using
11 20th century practices if they do not work
12 in the 21st century.

13 Which means I do support the CBT,
14 which is computer-based assessments that
15 many experts discussed today. I think it
16 would be beneficial for all of our students
17 to take part in a CBT. However, I think
18 that we need to have also other assessments
19 to go along with that CBT, computer-based
20 testing, such as writing.

21 We constantly -- right now we're
22 using CB testing, such as SATs, ACTs, and
23 other formative type assessment -- summative

1 type assessment right now to evaluate
2 students and to diagnose them. I think it
3 will be important if we can modify that same
4 structure that we use to have students enter
5 into college and even get into the career
6 fields if we can modify that so it could fit
7 for early childhood.

8 No one spoke about that -- having
9 students being ready for the career field
10 starting as early as kindergarten. We all
11 do agree that students who are prepared
12 traditionally do better on assessments, as
13 well as in school if they're prepared early
14 with a head start.

15 I do feel that Georgia is getting
16 it right. We are doing wonderful things
17 here in the state of Georgia for
18 kindergarten students in particular where we
19 are assessing them on all of the standards.

20 However, we need to continue to move a
21 couple of steps forward.

22 As we continue to design
23 computer-based tests with experts on this

1 panel, if we could adopt some of those
2 things here in the state of Georgia I think
3 we would continue to be a model for the
4 nation on how we can help students progress
5 through the educational parameters.

6 I think with federal, as well as
7 state, support many districts will be very
8 successful at rolling out new computer-based
9 type assessments. We must have teacher
10 perspective on these panels that also speak
11 to what the teacher actually experiences on
12 an everyday basis and what would be best for
13 students, which mean I think a balanced
14 testing approach would be very beneficial.

15 However, I do feel that
16 electronic portfolios will be something that
17 will be very progressive and innovative for
18 all of our students -- teachers, as well as
19 administrators, will be able to produce
20 yearly. Which means that if you could put
21 funding for some sort of way that we could
22 actually do more staff development and
23 training around creating electronic

1 portfolios for teachers, students, as well
2 as administrators, I think it will also be
3 very productive.

4 One person spoke about biases
5 today. I encourage us not to forget the
6 biases that come along with some of these
7 computer-based assessments, gender biases as
8 well as culture biases. Boys and girls do
9 learn differently and I think that's a very
10 important aspect that we must not forget as
11 we are designing different assessments.

12 We talked about several models
13 here today, and I think many of those models
14 have a lot of things that are working. If
15 we look at all of those different models and
16 pick what's best I'm sure that we in the
17 United States will have one of the best
18 models that many countries will really look
19 at.

20 And so as we have our structure
21 as far as our Race to the Top I'm sure that
22 we will continue to be leaders in the New
23 World with computer-based testing but also

1 great education that are giving our
2 educators and received by students. Thank
3 you.

4 MS. WEISS: Thanks so much.

5 DR. EADS: "Please speak directly
6 into mike." Thank you. Okay. That's the
7 sign down here.

8 I'm Jerry Eads. I am the -- I
9 work for the State of Georgia in the Teacher
10 Licensing Agency. I have to invoke the
11 Fabrizio disclaimer this morning, which is
12 I'm speaking for myself and not for them,
13 and this is primacy, recency. Hopefully I
14 do you justice, Lou.

15 Hi, Tom, by the way.

16 I want to speak on four things,
17 and I'll actually cover three, because the
18 other one's been touched on enough. One is
19 validity, the other is test design, the
20 third is testing time, and the final one is
21 society impact.

22 The Race to the Top assessment
23 executive summary notes that the framework

1 would focus on the design and quality of
2 assessment systems and not accountability
3 policies, yet it is the accountability
4 policies which drive validity, and validity
5 is at the very core of the issue of quality.

6 The testing system you want must
7 be validated for each of your purposes --
8 instructional improvement, measuring school
9 principal and teacher effectiveness, and
10 predicting college readiness -- whatever
11 that is -- for which college, as you pointed
12 out.

13 It would behoove us, for example,
14 to require consortia to demonstrate that the
15 system differentiates among teachers on
16 meaningful dimensions. Just because scores
17 go up does not mean in and of itself that
18 anything of value changes -- higher
19 graduation rates, increased college success,
20 lower unemployment rates, employment
21 persistence, et cetera. Unless we can
22 demonstrate externally referenced value the
23 testing system is nothing more than the

1 proverbial boat -- that is to say a hole in
2 the water into which we pour money.

3 Developing an assessment system
4 without considering policy intent and
5 implication is little different from
6 building an atom bomb and refusing to
7 address the consequences. Your framework
8 begins with policy validity issues, not the
9 least of which is individual student
10 achievement as measured against standards
11 that build toward college and career
12 readiness by the time of high school
13 completion.

14 We cannot know that the tests are
15 valid for that purpose unless we undertake
16 the requisite longitudinal work to determine
17 the relationship of the test scores to the
18 desired outcomes -- I urge you to require
19 such study.

20 Test design -- standards are
21 nothing more than minimum competency, so it
22 matters not whether the standards are set at
23 tenth, thirtieth, or ninetieth percentile

1 level. They are still just news speak for
2 minimum competency.

3 The executive summary talks about
4 standards and growth, the latter requiring
5 full range testing in the same breath. If
6 standards become, as they are now, nothing
7 more than different levels of test
8 performance then it will be hard to avoid
9 setting performance levels for different
10 goals -- college, tech school, direct job
11 placement -- on the same test similar to
12 what some European countries do. It's
13 called tracking.

14 We rarely consider the
15 consequences of so-called high standards,
16 dropout rates, narrow curriculum and so on,
17 and thus we rarely try to address those
18 problems at the policy level but leave the
19 classroom teacher to struggle in the
20 aftermath of our short-sightedness.

21 We have yet to establish that, in
22 fact, high stakes tests of any nature
23 actually improve education. I dearly

1 require you requires states to study the
2 long-term impact of testing policy. If
3 scores go up and teachers get hired and
4 teachers get fired and nothing else changes
5 then the millions spent on testing is for
6 naught.

7 We focus our testing on standards
8 in basic areas. It's been adequately
9 demonstrated that testing in only a few high
10 stakes areas narrows instruction. The
11 accountability movement seems to focus only
12 on the development of skills to produce good
13 workers with virtually no consideration for
14 other purposes of public education, such as
15 producing good citizens.

16 What is the impact of society --
17 on the society of reducing in the arts --
18 reducing the arts, language, and social
19 sciences? We pay lip service to a desire to
20 test critical thinking, yet we seem
21 determined to remove from schooling those
22 things we're thinking about.

23 Perhaps it would be useful to

1 undertake the study of the relationship
2 between test scores and indicators of
3 societal outcome such as incidents of voting
4 and arrest.

5 Teachers in public schools
6 serving upper socioeconomic strata spend
7 little time preparing students for tenth and
8 twentieth percentile standards. But if
9 teachers in poor schools must spend all
10 their time getting their students to pass
11 minimum competencies in English and math,
12 how does that impact instructional time in
13 other areas? And do those changes, if any,
14 have impact on such things as citizenship
15 behaviors?

16 I do indeed understand we must
17 start somewhere and that there are
18 reasonable arguments for beginning this
19 effort with reading and math. But let us
20 make sure that Mr. Obama's initiative does
21 not do to our public schools what someone
22 else did to Afghanistan: getting things off
23 to a reasonable start and then walking away.

1 Thank you.

2 MS. WEISS: Thank you. I take it
3 from the empty chairs up here that that's it
4 for our public speakers.

5 So let me just conclude by
6 thanking all of the people who just spoke
7 with us for taking their time and sharing
8 their thoughts with us.

9 And, most sincerely, I'd love to
10 thank now our experts who have been here
11 listening. I see them scattered throughout
12 the audience. So those of you who were
13 doing public speaking fear not, even though
14 the table up here is empty, they're
15 everywhere out there.

16 But I want to just thank our
17 experts for sharing their time with us, for
18 sharing their wisdom and their experience
19 and their expertise with us. I hope that
20 all of you learned something today that you
21 didn't know before you walked into this
22 room. I know that those of us up here from
23 the Department certainly did.

1 So we just want to thank
2 everybody most sincerely. I hope that the
3 states have benefitted from this activity.
4 I know we'll see many of you, but not all of
5 you, back here tomorrow. Tomorrow morning
6 we're doing a half-day sessions specifically
7 focused on issues of students with
8 disabilities.

9 So if you're coming tomorrow we
10 look forward to seeing you then. If not,
11 thank you so much for spending your day with
12 us today. We really appreciate it and hope
13 that it was as helpful for you as it was for
14 us. So thank you.

15 (Whereupon, at 4:15 p.m., the
16 meeting was recessed, to resume Wednesday,
17 November 18, 2009.)
18
19
20
21
22
23

1

2

3

4

5

6

7

8

9