

UNITED STATES DEPARTMENT OF EDUCATION

RACE TO THE TOP ASSESSMENT PROGRAM

PUBLIC & EXPERT INPUT MEETINGS

GENERAL & TECHNICAL ASSESSMENT

Washington, D.C.

Wednesday, January 20, 2010

1 PARTICIPANTS:

2 Panelists:

3 JOANNE WEISS, Moderator
4 U.S. Department of Education

5 JOHN EASTON
6 U.S. Department of Education

7 ANN WHALEN U.S. Department of Education

8 JUDY WURTZEL
9 U.S. Department of Education

10 Presenters:

11 JAMAL ABEDI
12 University of California - Davis, School of
13 Education

14 RANDY BENNETT
15 Educational Testing Service, Research and
16 Development

17 LIZANNE DeSTEFANO
18 University of Illinois, College of Education

19 SCOTT MARION
20 National Center for the Improvement of Educational
21 Assessment

22 JEFF NELLHAUS
Massachusetts Department of Elementary and
Secondary Education

LAURESS WISE
Human Resources Research Organization

Public Speakers:

LINDA DARLING-HAMMOND

JOHN FOSTER

1 PARTICIPANTS (CONT'D):

2 SUSAN GENDRON

3 NANCY GREEN

4 MYRTICE IRISH

5 BARBARA KAPINUS

6 CHRISTINE MASON

7 JOHN OLSON

8 RICHARD PATZ

9 BILL RING

10 ROSA VILLEGAS

11 WALTER RICHARDSON

12

13

14 * * * * *

15

16

17

18

19

20

21

22

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

2 (10:00 a.m.)

3 MS. WEISS: So good morning, everybody.
4 Thank you so much for joining us today on what we
5 absolutely, positively really do promise is the
6 last of these meetings that we will be subjecting
7 -- no, I mean, hosting -- for everyone. So thank
8 you for coming. I think we have a very
9 interesting day ahead of us.

10 And let me start by just talking for one
11 moment about that other competition. I know a lot
12 of the states who are here today got to hand in
13 their applications personally yesterday to us and
14 then stay for today and join us at this event.
15 So, we're glad you did that. We received
16 yesterday applications from 40 states and the
17 District of Columbia. So we got 41 applications
18 yesterday, exceeding, I think, all of our perhaps
19 wildest expectations. So that's fabulous.

20 This before you is the list of the
21 applications that we received yesterday, so thank
22 you for that. And lest we spend a minute letting

1 grass grow under our feet here, we are today
2 talking about the next set of issues.

3 So with that I'm going to turn to the
4 next set of issues. So you all know that we are
5 talking about the Race to the Top Assessment
6 Competition today designed to support consortia of
7 states in implementing sets of common standards by
8 funding the cost of developing new assessments to
9 support the standards. The timeline we've talked
10 about before. We're still operating on the same
11 timeline, which is that we are hoping to get the
12 notice out in March. Applications will be due in
13 June and grants awarded in September.

14 The goals of the program remain the same
15 as well. I'm going to run through them very
16 quickly and then we'll turn our attention to the
17 purpose of today's meeting. But the goals of the
18 program remain that we are trying to support
19 states in delivering a system of perhaps more
20 effective and instructionally useful assessments
21 than we have had in place to date. We want to
22 make sure that we get accurate information about

1 what students know and do, and that involves
2 achievement against standards; it involves student
3 growth; and it involves being able to give
4 students information about whether they're on
5 track to be in college and career ready by the
6 time of high school completion.

7 We also want to make sure that we are
8 developing tests that reflect and support and
9 promote good instructional practice and that from
10 the very beginning include all students, including
11 English language learners and students with
12 disabilities in the way that we're thinking about
13 conceiving and developing the tests.

14 We've also talked in past meetings about
15 the fact that we're at a slightly uncomfortable --
16 from a timing point of view, we're at a slightly
17 uncomfortable place with this, which is to say
18 that we're developing tests or putting out a
19 competition at a time that straddles the
20 regulations under No Child -- under No Child Left
21 Behind and whatever the new ESCA reauthorization
22 will look like. And so we are trying to develop a

1 set of tests that both comply with the existing
2 law, as well as giving us the freedom to think in
3 some new ways about some of the directions that we
4 might want to go. So the guidance that we've
5 given so far is that we want these assessments to
6 be useable to inform teaching, learning, and
7 program improvement; determinations of school
8 effectiveness for accountability purposes;
9 determinations of principal and teacher
10 effectiveness for both the purposes of evaluation
11 and support; and determinations at the individual
12 student level of college and career readiness.

13 So consistent with No Child Left Behind,
14 we do want to cover at a minimum reading, language
15 arts, and mathematics annually from grades 3
16 through 8, as well as at least once in high
17 school. We are focusing, although not necessarily
18 exclusively, but certainly these assessments --
19 the competition needs to produce a new set of
20 summative assessments, but we're not necessarily
21 talking about summative assessments that look like
22 the ones that we've got today. So they don't

1 necessarily have to be only end of the year tests.
2 They don't necessarily have to be given only once
3 during the year. It may not be one test; it may
4 be a system of tests that we need to accomplish
5 our different purposes. We are thinking though
6 that these new tests would replace rather than add
7 to the assessments that we've got out there
8 already.

9 So that's sort of the big picture
10 framework that we've been operating within for all
11 of these meetings. The reason for holding these
12 meetings is really threefold. First, we just felt
13 that it was going to be helpful for both us in the
14 Department, as well as for states who are
15 competing, to really have a vision of what the
16 next generation of assessments could and should
17 look like. That we all were so mired in the today
18 that we needed the opportunity to pick our heads
19 up and think of the possibilities, and we invited
20 experts to help us paint those pictures so that we
21 could start developing different visions of what
22 could be.

1 At the same time though, we also asked
2 our experts to provide very concrete guidance to
3 us in the Department of Education in response to
4 the questions that we asked in the notice so that
5 we could get very specific thoughts, words, that
6 we should be using in our notice to elicit the
7 right kinds of assessment designs to accomplish
8 the goals that we had set forth.

9 And the third thing that we wanted to do
10 is we know that really a key audience for these
11 meetings has been and continues to be the states
12 who are the applicants for this. And that we
13 really share with you the need to have a deeper,
14 better understanding of both a vision and what
15 could be. We have said numerous times that we
16 wish that you were able to be on stage with us.
17 Because you are applicants, we are not allowed to
18 do that, but nonetheless we're really delighted
19 that so many states have been able to come to
20 these meetings, participate in them, and hopefully
21 learn along with us.

22 So at this point we have heard input

1 from 42 experts and 79 members of the public over
2 the course of the meetings that we've held. We've
3 also gotten over 50 pieces of written input. All
4 of this information -- transcripts, PowerPoints,
5 written input -- are all available on the
6 Department's website, as will this meeting be a
7 couple of days from now. And so as we put final
8 pen to paper we thought we would need sort of a
9 summary meeting with some of the experts from each
10 of the panels that we've held to date to come in
11 and just comment one last time on whatever
12 lingering questions we had.

13 So about a week and a half ago or so we
14 published just a very brief sort of addendum to
15 the notice that we had put out that said, okay,
16 here's the final questions that we have that we
17 want to work up for today. It's actually -- I've
18 got them on slides that I'm just going to fly
19 through in a second. They're actually somewhat
20 long and complex questions so I have boiled them
21 down to the four high level talking points here.

22 The first one is that we are considering

1 this notion of doing what some have called
2 through-course summative assessments -- summative
3 assessments that are given periodically over the
4 course of a school year and rolled up into a final
5 summative score. And we've asked our panelists to
6 talk about whether this is a good idea. Is it
7 something that we should require? Encourage? Be
8 silent on? And why? And also some of the big
9 questions that came up to us as we were looking at
10 this were questions about validity and reliability
11 for accountability purposes. These summative
12 assessments do have to carry accountability weight
13 and we wanted to make sure that a new design like
14 this -- it's not necessarily new to the world
15 internationally. It's a model that's used in many
16 places, but it's certainly new in this country.
17 Did we all feel like we had the tools that we
18 needed psychometrically to be comfortable with
19 something like this? So that's sort of the first
20 big question.

21 The second big question is that we're
22 also thinking about high school end-of-course

1 assessments as a way of really promoting increased
2 levels of rigor, increased equity across high
3 schools to really rigorous high school work and
4 courses, and a model of just increasing
5 instructional quality, large at the high school
6 level which I think we all know is desperately in
7 need of improvement. And we wanted to understand
8 whether some of the ideas we had put forward would
9 really help us achieve those goals. And was it
10 possible and how could we think about achieving
11 levels of consistent high rigor in a system that's
12 pretty, the way we've been thinking about it,
13 decentralized run by states or by consortia and
14 not necessarily part of at least a federal
15 accountability system.

16 The third question we had was about uses
17 of technology. And the biggest question we had
18 there was that based on a lot of the input that
19 we've heard from people, it seems like if we're
20 putting out an assessment that's designed to be --
21 that is designed to go into wide scale
22 implementation four years from now, that requiring

1 technology as the base model of delivery with
2 paper and pencil perhaps as an accommodation
3 rather than the other way around, might be the
4 right way to go. And we wanted to just test that
5 with folks one more time up here.

6 And then the last question -- it's
7 actually the last two questions in the notice that
8 we put out -- are around the need for innovation
9 and the need for additional research. Just the
10 issues of balancing the actual requirements we
11 have to have tests come out of this at the other
12 end of this four-year period with how much we need
13 to do additional investment in innovation or in
14 research, and whether that's the federal
15 government's job or whether there's other ways to
16 accomplish those goals outside of this grant.

17 So those are the big questions.
18 They're, as you can tell, big questions that we
19 ask people to address in, oh, 20 minutes. So we
20 will turn it over to them in a moment and get
21 started. These, as I promised, are the big
22 questions that you can't even read from where

1 you're sitting, I'm sure, but they're in the
2 notice if you'd like to see them.

3 The schedule for today is that we'll
4 have four presentations before lunch. Each
5 presentation will be 20 minutes long followed by
6 10 minutes of clarifying questions from any
7 members of the panel up here. After lunch we'll
8 have two more presentations followed by a
9 roundtable discussion. And after that, for the
10 last hour of the day, we will hear from a number
11 of people who signed up ahead of time to be public
12 speakers today.

13 I think all of you got note cards when
14 you registered. Those are for questions. Feel
15 free to submit questions when you've got them.
16 You can give them to any members of the Department
17 staff or drop them at the table out there and
18 they'll get the questions to us. And we will do
19 our best to feed them into the conversation that
20 we're having. We will be doing pretty rigorous
21 timekeeping up here, so if you hear us talking
22 about how much time we all have left, it's because

1 there's timers blinking at us as we talk. Please
2 do put your cell phones on vibrate. And just one
3 more additional note that we will be accepting
4 written input through the end of the day today.
5 I've put the address up here to e-mail written
6 input to and everything that we've done so far is
7 posted on our website, ed.gov.

8 A quick thank you to all the states who
9 are attending today. We are also doing this by
10 webinar, so all the states in italics are on the
11 webinar listening in to everything that we're
12 talking about today. And the ones in plain type
13 are sitting here in the room with us, so thank you
14 all for coming today. And with that let me just
15 go around the panel and ask each person to
16 introduce themselves. And then we'll get started.

17 Laurie?

18 MR. WISE: I'm Laurie Wise. I'm with
19 the Human Resources --

20 MS. WEISS: And pull your mics up to
21 you, if you don't mind, just because otherwise the
22 people listening in on the webinar can't hear.

1 MR. WISE: Hello, webinar. I'm Laurie
2 Wise. I'm with the Human Resources Research
3 Organization, or HumRRO. And I'm advisor to
4 several states and work on a number of projects
5 having to do with test and test use policy.

6 MR. ABEDI: Jamal Abedi, University of
7 California, Los Angeles -- Davis. It's still in
8 Los Angeles. I have -- my research focus is on
9 assessment issues for (inaudible) learners and
10 I've advised several states on their assessment
11 issues.

12 MS. DeSTEFANO: My name is Lizanne
13 DeStefano. I'm a professor of educational
14 psychology at University of Illinois in
15 Urbana-Champaign. My interest is evaluating local
16 implementation of federal initiatives and also the
17 impact of assessment and accountability
18 initiatives on special populations, like English
19 language learners, students with disabilities,
20 students of poverty, and so on.

21 MR. NELLHAUS: Good morning. I'm Jeff
22 Nellhaus. And I just was promoted this morning.

1 I see I'm listed as commissioner of education for
2 the Massachusetts Department of Elementary and
3 Secondary Education. In fact, I'm the deputy
4 commissioner in Massachusetts. And --

5 MS. WEISS: A round of applause for Jeff
6 and his new position.

7 MR. BENNETT: I'm Randy Bennett from
8 Educational Testing Service, and my research
9 focuses on integrating advances in cognitive
10 science, measurement, and technology to create
11 models for new approaches to assessment.

12 MR. MARION: And I'm Scott Marion. I'm
13 the associate director of the National Center for
14 the Improvement of Educational Assessment. We
15 work with actually over half the states on a range
16 of assessment and accountability issues.

17 MS. WEISS: And I'm Joanne Weiss. I'm
18 the director of the Race to the Top program at the
19 Department of Education.

20 MR. EASTON: My name is John Easton.
21 I'm the director of the Institute of Education
22 Sciences at the U.S. Department of Education.

1 MS. WURTZEL: I'm Judy Wurtzel. I'm the
2 deputy assistant secretary for planning,
3 evaluation, and policy development at the
4 Department.

5 MS. WHALEN: And I'm Ann Whalen with the
6 Office of the Secretary in the Department of
7 Education.

8 MS. WEISS: And with that I'm going to
9 hand the baton over to Scott and we'll get going.

10 MR. MARION: Thanks, Joanne. And thanks
11 again for inviting me. I appreciate the
12 opportunity to speak. And knowing that I have 20
13 minutes I'm going to go through some of these
14 slides rather quickly, and others where I want to
15 emphasize points I'll linger a little bit longer.

16 So, in addition to the five questions --
17 for those of us who are Jewish in the audience,
18 it's like Passover plus one. It's a lot in a
19 short amount of time. But I have some
20 introductory comments first. And I think some of
21 these are some of the homework assignments for my
22 Department colleagues here. But some things to

1 think about as framing the proposals.

2 So, the questions that we received and
3 other written documents that have come from the
4 Department, whether in the Federal Register or in
5 other forms, either imply or directly appear to
6 propose to require a specific way of doing things.
7 And I would argue unless you are absolutely sure
8 this is the only or the best way to do things for
9 all context of accomplishing the goals that you
10 want to accomplish, then I would avoid requiring
11 the specific means and really focus on defining
12 the ends, to be exceptionally clear about the
13 goals of the proposed systems. So not just to
14 have, for instance, a system that is common across
15 all the states because then the next question is
16 why. What goals is that going to further? The
17 purposes and uses -- beating a dead horse. I
18 mean, I know that the reauthorization will move
19 through Congress eventually. And that defines
20 some of the purpose and uses, but to the extent
21 that we could be clear now, it would be important
22 to do so.

1 And then if you could specify these
2 goals very clearly and at least describe the
3 purpose and uses, you could have a lot of the
4 smart proposal writers be creative and innovative
5 about the specific means. And if you're vague in
6 the notice, then the writers will probably have to
7 be even more vague and you lower the chances of
8 actually getting what you want.

9 This is the sort of tough love, tough
10 choices, right? As you read into this, what we're
11 asking for is innovation, broad implementation,
12 and fast timeline, or in multiple choice lingo,
13 all of the above. And something will give.
14 Everything we know about wide-scale
15 implementation, something is going to have to
16 give.

17 Lorrie Shepard, I thought, gave very
18 good advice at the Denver meeting in December.
19 And said allow consortia to propose to do one
20 relatively small thing, for instance, create an
21 innovative assessment system for grades 4 to 8 in
22 mathematics for just a few states, to at least

1 have a model of what might be able to be done and
2 done well if innovation is the goal. If broad
3 implementation without innovation is the goal,
4 then that leads you down a different path. But
5 it's important to make these choices.

6 And again, when you think about these
7 tough choices and this ends means discussion, you
8 get where these two actually run right into each
9 other. So question 2 as I'll talk about it in a
10 little more detail, asks about increasing the
11 rigor and quality of high school assessments.
12 Question 3, asks about the potential of requiring
13 computer-based testing. Requiring two might
14 hinder three, or requiring three might hinder two.
15 And so it's important to be really clear again
16 about what we're trying to accomplish.

17 I talked enough about a theory of action
18 in Boston in November. Bottom line is the theory
19 of action will allow -- especially in the
20 proposals -- will allow for a check of the logic
21 of the underlying assumptions being put forth in
22 the proposals.

1 One of the other things as I was
2 thinking about this is the operational
3 requirements of any of these multistate consortia
4 are overwhelming. And no state or set of states
5 has the capacity to design, implement, field test
6 this kind of multistate assessment system.
7 Consortia will have to issue RFPs to support the
8 design, development, and eventually the
9 implementation. But I know this doesn't carry us
10 through implementation. And so I thought, well,
11 maybe if the actual response was written as a RFP
12 it would allow the proposal writers -- a
13 well-written RFP will make the goals' rationale
14 design clear to potential bidders. It will also
15 reveal to the Department the extent to which the
16 proposers have thought through many aspects of the
17 proposed assessment system. As folks like Jeff,
18 who write a lot of RFPs know, the more clear you
19 are in your RFPs, the better you'll get in terms
20 of bids and likely contracts.

21 So, the Department has managed very
22 skillfully to avoid any mention of curriculum and

1 instruction in this set of questions. And I
2 understand the political necessity of that. But
3 it doesn't really make any practical conceptual
4 sense if the goal is really to move our
5 educational system forward. So, while you can't
6 mandate curriculum -- I understand that -- I at
7 least would recommend requiring all proposals to
8 at least address how their assessment model deals
9 with these considerable differences in curriculum
10 instruction across districts and states. How do
11 the proposers think these differences will affect
12 the assessment results? How do they propose to
13 deal with these differences, if at all? And how
14 do they think their assessment model will be able
15 to further meaningful goals if they did not deal
16 with curriculum instruction. I'd love to see that
17 to be required in the proposal.

18 So, I didn't repeat all the questions
19 here. They're in the notice. I suspect many of
20 you have downloaded them. If you had really good
21 eyes you could read them on Joanne's slide.

22 Question 1 was asking about this

1 through-course approach. For each of these I'm
2 going to try to give a few quick comments and then
3 some recommendations for evidence. So, I find a
4 lot of aspects of this approach appealing. I
5 mean, I proposed something not dissimilar to this
6 in November. But again, I would refrain -- as
7 much as I sort of resonate with this, I would not
8 require a specific approach unless I knew exactly
9 what the goals were and this was the only approach
10 to meet these goals.

11 So what are we or the U.S. Department
12 (inaudible) consortia trying to accomplish with
13 the through-course approach? All the proposals,
14 whether they're using a through-course approach or
15 not, should be required to submit evidence or at
16 least a rationale in all six categories outlined
17 in the question. The through-course approach,
18 however, carries some unique considerations for
19 evidence. And since it was part of the question I
20 felt the need to address it.

21 Inter-rater reliability is not one of
22 those additional concerns with this approach.

1 People obsess about inter-rater reliability. And
2 I think it's only because they're able to compute
3 it easily and they can show something
4 quantitative. It's the least of our worries in
5 this kind of system. It's not that it's not
6 something to be concerned about, but it's way down
7 the list. And most vendors have really good
8 systems with monitoring that.

9 But a through-course approach proposal
10 should have to provide evidence for -- like all
11 approaches -- construct validity. But how would
12 this approach enhance the validity of the scoring
13 interpretations compared to the single end-of-
14 year test? Aggregation is a huge issue. We know
15 very little other than trying to maximize
16 reliability about how to test aggregate the scores
17 for multiple assessment events for both students
18 and schools in order to maximize validity or
19 maximize utility. How will the states and
20 consortia deal with potential increased effects
21 due to opportunity to learn differences and
22 opportunity to learn?

1 Security. If these, in fact, are used
2 for accountability, it raises some considerable
3 security concerns with these through-course
4 components.

5 The consequences. How will the
6 consortium deal with the potential and perhaps
7 likely negative effects when educators are
8 restricted from using the full potential of these
9 through-course components for instructional
10 improvement. What I was proposing -- this more
11 comprehensive system -- don't get me wrong. I'm
12 fully in favor of a comprehensive system, but
13 that's because I want to be able to use the stuff
14 that's going on on a day-to-day, or week-to-week,
15 or month-to-month basis for instructional
16 improvement. If this is in a secure environment
17 it lessens the opportunity to use these things for
18 instructional improvement.

19 I don't think we do equating that well
20 now in terms of monitoring changes of scores
21 across the years with different assessments.
22 Doing these kind of -- these memorable events

1 through the course will add to our equating
2 challenges. Not that we can't deal with them, but
3 I want to know how the consortium proposes to deal
4 with them.

5 High school rigor. I'm thrilled that
6 the Department is thinking about increasing the
7 quality and rigor in high school. I'm not sure
8 why it has to be a common end- of-course summative
9 examine considering the previous question as well.
10 So, you know, why not think about it as a through-
11 course? But I want to know what the unit is for
12 common. Is it the school, the district, the
13 state, or the consortium? I'm hoping it doesn't
14 have to be the consortium. I would argue, too,
15 there's mention of the words "consistent" or
16 "common" in the question. I think it's much more
17 important to focus on rigor because we do that
18 less well. We're pretty good at consistent and
19 common.

20 My colleague Brian Gong and I in another
21 context have written about these tradeoffs between
22 flexibility and standardization. And it's worth

1 attending to that here because if we got for
2 standardization we could certainly lose innovation
3 and rigor.

4 Amy Gutmann of Loss of Education, who is
5 actually now I think the dean at Princeton, has
6 written about this conception of a threshold. And
7 so not that everybody has to be the same, but we
8 want to make sure children can get past a certain
9 threshold of expectation. Look, we don't require
10 all doctoral dissertations to be fully comparable,
11 but we give people degrees that have some sort of
12 similarity of threshold.

13 The type of evidence I would require for
14 this rigor -- I would require evidence that
15 student performance meets a meaningful threshold.
16 I would want to know what the system of review
17 within the consortium for rigor and technical
18 quality and how does the consortium propose to
19 make this work within and across states. Has the
20 consortium addressed the balance convincingly
21 between standardization and flexibility and
22 offered a good rationale for where they stand? I

1 think it would be really important if the
2 Department could signal clearly where they think
3 is the right balance between standardization of
4 flexibility to help the states in their proposals.
5 And how will the consortia ensure that students
6 have a fair opportunity to meet these rigorous
7 thresholds? Validity is threatened if opportunity
8 to learn is not provided.

9 So sitting next to Randy Bennett I
10 hesitate to even talk about this computerized
11 testing stuff, but I'll try. Again, I wouldn't
12 require -- I know this will be my mantra today --
13 I would -- tell me what you want to accomplish
14 with this and see if this is the best way to do
15 it. If we want innovation and we want the kind of
16 innovation that we can get with computers I would
17 not, again, just like I said I wouldn't focus on
18 common inconsistency in the previous; I wouldn't
19 focus too much on comparability. I know from
20 states that I work in we can't even -- if you want
21 to really be strict about comparability, we can't
22 even ensure comparability of computer-to-computer

1 comparability within a single state. We've been
2 learning a lot about differences in resolution.

3 Screen sizers, processors, it's mindboggling. I
4 actually think within computer variability is
5 greater than the between computer and paper
6 variability in certain states. So you could
7 obsess about it, but then you would hinder
8 innovation.

9 I think that CBT offers considerable
10 potential for enhancing access for students with
11 disability, but it could also increase a construct
12 of relevant variance. We have to balance these
13 two. There's a group out of Boston, Nimble Tools,
14 but they're just one. There are others that have
15 demonstrated the potential of doing this well --
16 of building it in in terms of access.

17 Again, the evidence to support CBT, I
18 would want to know the items are designed or at
19 least working towards design to take advantage of
20 the technological capabilities. It shouldn't be
21 just a way of translating items from paper to
22 computer. How will the consortium states move to

1 full implementation of CBT so they can begin using
2 innovative item types? That's where you could run
3 the risk of comparability. If you have these
4 really cool interactive items that you can only do
5 on the computer, it's hard to find a paper analog
6 to that. And so I've learned a lot from the
7 states in which we work that getting to 70 percent
8 implementation is actually doable. It's sort of
9 like golf. You know, if you can get to 80 you're
10 doing really well, but getting to 70 is really,
11 really hard. And it could take a lifetime, or
12 then some.

13 And so getting that last group of
14 schools that for whatever reason can't do it is
15 tough. What type of design-in -- not add-on --
16 approach is the consortium approaching for
17 increasing access for students with disability and
18 the ELLs? And how will -- this is something I
19 learned about just last week -- how will the
20 consortium states avoid the negative consequences
21 of the loss of computers and computer time for
22 instructional purposes? If the labs -- especially

1 in high schools -- are being used for testing,
2 especially on a through-course approach, then
3 you're going to have labs not available for
4 instructional purposes.

5 Beyond this people are talking about
6 computer adaptive testing. And there's lots of
7 issues here in terms of item bank size and things
8 like that. One thing the Department will have to
9 wrestle with -- you know, don't anybody shoot me
10 for mentioning these three words together -- but
11 how will out-of-grade level items be allowed? If
12 not, the potential of CAT is limited considerably
13 at least for one purpose. But, of course, this
14 must be balanced with social justice concerns.

15 This is about the grant timeline and
16 innovation. Again, I would not encourage or fund
17 grants that do not move down a path towards
18 innovation. Randy was really eloquent in this in
19 Boston in November. There's a few things I want
20 to pick up on. States and consortias clearly need
21 to articulate a vision of what they hope to
22 accomplish with their educational system in 10

1 years or so and provide evidence or justification
2 for how the proposed assessment system will
3 support this vision. The assessment system is not
4 an end in and of itself. So you want to see a map
5 -- not just a map though -- but a map and a route.
6 And how you're going to get there. A theory of
7 action that describes how the states and consortia
8 will be able to stay on that route will also be an
9 important piece of the proposal.

10 And I'd want to see evidence that states
11 and consortia have at least thought about the idea
12 that they're not going to paint themselves into a
13 corner so that in four years at the end of this
14 grant period they're no further along, perhaps
15 even further away from, this 10-year vision than
16 they were when they started. This is another
17 reason in my mind for funding multiple consortia
18 that could tackle manageable-size programs.

19 This is getting sensitive. In terms of
20 research proprieties -- I'm glad John is on the
21 panel today; it was nice to see that -- I think --
22 and my organization does a lot of work -- Damian

1 Betebenner's student growth percentiles -- people
2 liken it to them. He would argue it's not, but
3 it's close enough. I think the statistical
4 machinery of VAM has been well studied. And we
5 don't need special funding for this. More
6 research is not going to correct for nonrandom
7 assignment. I mean, we're working hard to do this
8 one thing and we could do it about as well as we
9 can. And so we have to then be smart about the
10 uses.

11 As a little advertisement, the National
12 Research Council is coming out with a document
13 about -- a study panel put together the study
14 issues around evaluated modeling uses in
15 education. Actually, it should have been released
16 now, but it's coming soon. And that actually
17 summarizes a lot of these issues.

18 Related areas to sticking on this VAM
19 where we do need more funding -- the design and
20 validity of learning progression to support both
21 formative assessment and measuring growth with
22 summative assessments so that we actually can say

1 something meaningful about the growth that kids
2 have, not just that they added a few points or
3 that the school residual was positive.

4 We need assessment designs that allow
5 for meaningful depictions of student progress,
6 particularly related to learning progressions.
7 And we also need more research about how to
8 improve the quality and usefulness of VAM and
9 growth results for instructional improvements and
10 for accountability. We need more information
11 about how to integrate VAM results with other
12 observational evidence to make valid judgments
13 about educator quality. We know that we can't do
14 it quantitatively alone so how do we integrate
15 these sources of information?

16 And we have to figure out how to better
17 deal with these attribution challenges. So who
18 gets credit for the kids' success or who gets the
19 blame for the kids' failure particularly once you
20 get out of -- past 5th grade?

21 This thing is -- there we go. Research.
22 Question 2 was asking about the generalizability

1 and comparability of performance estimates.
2 Lizanne was involved in a lot of this. We learned
3 a lot about the generalizability performance
4 assessments during the 1990s. We could certainly
5 -- people are always happy to do more studies, but
6 that's probably been well studied. If I had
7 unlimited pie, we could certainly stand to learn a
8 lot more about how to integrate performance
9 assessments scores, especially if they're given at
10 a different time of year with a range of summative
11 assessment-type scores. We could stand to learn a
12 lot more about equating designs with performance
13 assessments or mixed assessment-types. If we
14 could learn to do equating well, then we could
15 more readily include performance assessments as
16 part of growth measures. That's really a
17 limitation right now. And we need better design
18 specifications and requirements for rich and
19 engaging performance tasks.

20 Some additional priorities in my last 45
21 seconds I'd like to say. So, you know, we're in
22 this sort of external accountability system

1 mindset. Do they work to actually achieve the
2 policy priorities or do other forms of school
3 reform work better? It would be nice to actually
4 do some studies before the reauthorization, but,
5 you know, there's been some work there.

6 I think equating is a considerable
7 challenge. People sort of hide it in the
8 machinery or IRT and other things, but equating
9 test scores when so much is changing is really
10 quite challenging. How do we go about validating
11 these college ready measures? How do we know when
12 we've reached good enough? And I think learning
13 progressions are a fruitful area for both
14 implementation and progress and it will require
15 massive development and validation program.

16 And those of you who doubted I could do
17 that in 20 minutes.

18 MS. WEISS: Impressive.

19 SPEAKER: Good job.

20 MS. WEISS: And we have no questions.

21 Next? No. Questions?

22 SPEAKER: Who is next?

1 MS. WEISS: Questions? Oh, now
2 everybody is flipping through looking up all the
3 questions they had.

4 MS. WHALEN: So when you were talking to
5 the question about innovation you talked about
6 having states describe or provide evidence how
7 they wouldn't paint themselves into the corner.
8 Can you expand a little bit on this? Because it's
9 hard to have people articulate both short-term
10 innovation, as well as kind of more long-term
11 innovation and how you balance a portfolio of risk
12 in an application like this. Do you have any
13 ideas how that could be articulated?

14 MR. MARION: Well, we sort of -- we do
15 know where we're at now in terms of our current
16 capacity and sort of current designs. So, for
17 instance, in the simplest case, if a consortium
18 proposed to just do sort of the same kind of thing
19 that we're doing now, but just involve more
20 states, I'd want to know the theory of action why
21 now being able to do some cross state comparisons
22 is going to further the goals I have of perhaps --

1 and this is where it's important to be clear about
2 the goals. But if my goal was to improve or
3 increase the number of kids who were truly college
4 ready or particularly for kids in poor minority
5 backgrounds, then I would want to know how simply
6 finding out that Massachusetts scores higher than
7 State X is really going to further that goal. And
8 so then you're left with a system that looks just
9 like our current system, but all it does is
10 include more states. And to me, that perhaps is
11 not painting us into a corner per se, but four
12 years from now we're still on the same starting
13 line. And so we haven't advanced the field.

14 So I think states could say this is --
15 my goal as a state is to really improve the
16 quality and rigor of the preparedness of kids
17 coming out of high school because of the global
18 economy -- whatever you want to talk about. And
19 this is what we're going to do with our education
20 system. This is how I think that an assessment
21 system could support that. And then I would want
22 to see -- this is part of this sort of mountain

1 making. If this is where I want to end up, what
2 steps now am I going to take to get there? And
3 how do I know, or what's my rationalities, that
4 this is a step that's going to move me further
5 down the line than another step that might take me
6 off the line. And so I'm asking states to do
7 that, but I think it starts with being clear about
8 where you want to go. That's really the most
9 important thing.

10 I think Randy has a lot of presentations
11 around this that really talk about where we really
12 want to go. If I wanted to go with any variety of
13 systems, whether it was technologically-based or
14 very intensive performance-based, I'd want to know
15 what am I going to do now. So if I wanted to
16 incorporate a lot more of my system in the future
17 is to incorporate a lot more performance
18 assessments then I could do much more rapid
19 scoring in terms of automation or these
20 interactive computer-type things like you see with
21 medical education, I would want to see the system
22 I'm designing now at least provide some --

1 (inaudible) some pieces of that and allows me to
2 get better at designing innovative item types and
3 to try those out and see how they work in
4 large-scale situations for particular purposes and
5 uses.

6 Is that -- I mean, it's -- the first
7 thing is to be very clear about where you want to
8 go. That's the key thing I would ask. Because
9 you're never going to get there if you don't know
10 where you want to go. And then to say how these
11 steps are going to further that. But I'm hearing
12 a lot of conventional talk. It might not set us
13 back, but it's not going to move us forward.

14 MS. WHALEN: From us?

15 MR. MARION: No. Some from you.

16 MS. WEISS: So, I have a question. And
17 maybe we can even back up to your slide 10.

18 MR. MARION: I don't know if I can do
19 that.

20 MS. WEISS: If the clicker will
21 cooperate with you. So while you're trying to do
22 that.

1 So this is the slide where you talked
2 about what we needed additional evidence for.

3 MR. MARION: Okay.

4 MS. WEISS: And you talked about
5 construct validity. And I'm wondering -- what I'm
6 trying to understand -- boy, it's just not --

7 MR. MARION: It's not that sensitive.
8 It might need a new battery. So, Mark, if --
9 before somebody else. There we go. There you go.

10 MS. WEISS: What I'm trying -- this one.
11 What I'm trying to understand is we could ask
12 billions of questions in this proposal -- in this
13 -- for the application, but assuming we want to
14 have some limit on that, are these the questions
15 you think are the most important questions around
16 -- like, reliability is not there. Construct
17 validity is there, but other types of validity
18 aren't. Comparability is not there. You talked
19 about later why you didn't put that on here I
20 think.

21 MR. MARION: Right.

22 MS. WEISS: I just want to make sure

1 that I understand, like, is this your
2 recommendation about these are the -- whatever it
3 is -- seven things that are -- six things that you
4 think are the ones that we should ask about? Or
5 are these just six things you wanted to tell us
6 more about?

7 MR. MARION: There are six things I
8 wanted to tell you more about. I think that the
9 criteria that you guys outlined in the question --
10 you do have comparability in there. I just think
11 that these are more important. And some could say
12 that all validity is construct validity. So it's
13 in there.

14 MS. WEISS: Okay.

15 MR. MARION: And things like
16 consequences are certainly part of validity.
17 Doing equating well -- I mean, it's all validity.
18 That's the bottom line. And it really is. How
19 you aggregate these scores affects the score
20 interpretations, which is really -- but these are
21 things that we're not that good at yet. And this
22 is why I want to know how consortium would deal

1 with these things.

2 This issue of -- perhaps, but I don't
3 know -- perhaps the worst thing you could do is if
4 you did these through-course assessments -- simply
5 just add up all the scores and divide by the
6 number of possible points. Right?

7 MS. WEISS: Right.

8 MR. MARION: I mean, that would be a
9 terrible thing to do. And so there's better ways
10 to think about aggregating to maximize validity as
11 opposed to maximizing reliability. We're good at
12 that part.

13 MS. WEISS: Oh, so you were aggregating
14 -- okay.

15 MR. MARION: So aggregation is an issue
16 of how you aggregate these through-course events
17 to come up with a determination of whether or not
18 the kid passed the class, is college ready,
19 whether the school gets credit for the kid being
20 in, you know, in the right group or the wrong
21 group. Things like that.

22 So these are just things that I would

1 want to see addressed in addition to the ones --

2 MS. WEISS: Okay.

3 MR. MARION: -- and perhaps more
4 emphatically than the ones -- yeah, I think those
5 are the more important ones.

6 MS. WEISS: Okay.

7 MS. WURTZEL: So, Scott, in one of your
8 remarks you suggested that we might ask states to
9 include an RFP or a draft RFP as part of their
10 proposal. So what would you expect would be the
11 kind of information evidence that we would see
12 from that that we wouldn't see from a typical
13 proposal that we would get? And what benefit
14 would you see that having for the states that were
15 going through the application process, if any?

16 MR. MARION: Yeah, this was one of my
17 trial balloons. And I had this thought and I
18 would look forward to in our more open discussion,
19 especially with folks like Jeff and Laurie that
20 respond to a lot of our RFPs. But I thought
21 actually that this -- and we've been talking about
22 this at our organization -- the amount of money

1 and the amount of time required -- I think
2 actually 350 is a lot of money depending on how
3 you spend it. It could go fast if you do the
4 wrong things. But I think they're going to need
5 help from vendors and others. So how do you
6 articulate that need for help? People do it as a
7 request for proposals.

8 So if the submission was an actual --
9 and I'm not actually saying as an addendum. I'm
10 saying the submissions should be an RFP with
11 perhaps some technical appendices to say why I'm
12 going to do this approach. How we're going -- I'm
13 not talking about the consortium management part
14 of it. That's a different aspect of it. But how
15 do I expect a vendor to think about aggregating
16 scores -- if it's a through-course approach --
17 think about aggregating scores and what's their
18 best proposal? These are the things as a
19 consortium we're concerned about. These are the
20 kinds of issues where we'd like input. And so at
21 least I could then see how states and consortia --
22 because if they -- if the states and consortia

1 write this proposal then they're just basically
2 going to the store and saying this is exactly what
3 we want to the vendors sitting out here and others
4 and say who's going to give me the lowest price.
5 And without allowing vendors to then engage in
6 proposing perhaps even more creative solutions to
7 this.

8 Like I said, I'm not entirely wedded to
9 this idea. I just think it might be a way to help
10 move us further and not lock us into too many
11 boxes too early. And so I would be interested in
12 hearing other people's thoughts about that.

13 MS. WEISS: Great. We'll come back to
14 that. Any other questions?

15 Okay, let's move on to Randy. You have
16 to click forward a bunch to get to yours.

17 MR. BENNETT: Okay.

18 MS. WEISS: Functioning battery.

19 MR. BENNETT: Okay. Thanks very much.

20 It's a pleasure to be here.

21 Question 1 concerns, as we've been
22 talking about, through-course summative

1 assessment. And I have three recommendations in
2 response to it.

3 The first recommendation is that ED
4 suggests considering through-course summative
5 assessment as a preferred rather than a required
6 model. And the reason is that there may be other
7 equally promising models. And ideally, I would
8 hope to see several consortia funded each
9 following a different model thereby allowing, in
10 essence, a real world trial of the viability and
11 effectiveness of those competing models.

12 My second recommendation is that at a
13 minimum ED requests the following evidence for
14 consortia proposing through-course summative
15 assessment or really any assessment for
16 accountability purposes. And as Scott suggested,
17 a theory of action certainly, but also a research
18 plan for evaluating that theory of action.

19 And I want to say a bit about each of
20 those in detail because I really think these
21 should comprise a significant proportion of any
22 proposal. The theory of action should name the

1 elements of the through-course assessment system.
2 For example, periodic tests, but perhaps also
3 project work and portfolios. The theories should
4 give a logical and coherent rationale for each of
5 those elements, including backing for that
6 rationale and research if that's available.

7 For example, for periodic tests, part of
8 a rationale might be that periodic tests are
9 intended to provide more timely feedback on
10 student achievement of standards than a single
11 end-of-year assessment would. And when
12 accumulated, give a more complete picture of
13 student accomplishment than that single
14 end-of-course assessment would. If project work
15 were an element to the beginning of a rationale,
16 it might be that project work is intended to allow
17 assessment of competencies that can't be measured
18 through periodic tests.

19 It should also be obvious that the
20 theory of action include the claims that will be
21 made from assessment results. So, for example, a
22 claim might be that student performance on

1 periodic tests and project work represents
2 achievement of common standards. Or it might be
3 that students who perform at the proficient level
4 -- whatever that is -- are ready to proceed to the
5 next grade's work. Or it might be that teachers
6 of classes with lower than expected performance
7 should be administratively reviewed because
8 they're likely to be ineffective.

9 Third, the theory of action should
10 enumerate the intended effects of the assessment
11 system and the mechanisms thought to cause those
12 effects. So, for example, that project work will
13 encourage a focus on important competencies not
14 promoted by traditional assessments; or that
15 linking teacher sanctions to student performance
16 will cause improved teaching practice; or that the
17 two of those in combination will lead to higher
18 achievement.

19 The second part of the recommended
20 evidence base is a proposed program of research to
21 evaluate that theory. A first question concerns
22 whether the theory is logical, coherent, and

1 scientifically defensible. That question might be
2 addressed through such mechanisms as TAC review
3 and early public presentation, invited critique by
4 independent experts, rejoined by the proponents,
5 and publication of the proceedings. The idea is
6 to expose the theory to the field early on so that
7 poorly considered ideas are quickly abandoned and
8 good ideas are vigorously pursued and improved.

9 Next, the program of research should
10 indicate how the consortium plans to evaluate the
11 stated assessment claims. For example, if a
12 consortium was claiming that student performance
13 on periodic tests represents achievement of common
14 standards, then I'd expect to see proposed an
15 alignment study and cognitive interviews to verify
16 that the processes students actually use during
17 assessment are indeed the intended ones. If a
18 claim was that students who perform at the
19 proficient level are ready to proceed to the next
20 grade's work, then I'd expect to see proposed a
21 predictive study to see, in fact, how students
22 actually perform in the next year.

1 If a claim was that teachers of classes
2 with lower than expected performance should be
3 administratively reviewed, then I might expect to
4 see proposed a blind observational study comparing
5 teaching practice in lower than expected and
6 higher than expected classes.

7 Third, the research program should
8 indicate how the consortium plans to evaluate the
9 implicit assessment claims, which typically are
10 necessary for satisfying the stated ones. For a
11 third course assessment system, the implicit
12 claims might include that scores aggregated across
13 periodic assessments can be compared; that
14 aggregated scores can be used to measure growth;
15 that scores from constructed response tasks,
16 including project work, are generalizable across
17 raters; and so on. These are very typical claims.

18 Last, the research program should
19 indicate how the consortium plans to evaluate
20 whether the assessment system was implemented as
21 intended; whether the intended effects were
22 achieved; and whether the postulated mechanisms

1 appeared to cause those effects. Because the data
2 needed to support strong causal claims will be
3 very difficult to collect under most
4 circumstances, the idea here is to gather whatever
5 evidence can be practically assembled to see if
6 it's consistent with the theory of action. And if
7 not, to try to identify why not.

8 So, for example, if an intended effect
9 was that project work would encourage students and
10 teachers to focus on important competencies, then
11 I might expect to see proposed a study of
12 classroom processes before and after advent of the
13 assessment system. If the intended effect was
14 that linking teacher sanctions to student
15 performance will cause improved teaching practice,
16 then I might expect to see proposed an analysis of
17 teacher lesson plans before and after the
18 assessment system was introduced.

19 My third recommendation concerns the
20 sequence and timing of through-course summative
21 assessment. To minimize such timing and sequence
22 effects, ED might strongly encourage each

1 consortium to agree upon an administration
2 sequence and set of administration windows and
3 provide a plan for protecting test content so
4 students taking the test later in the window don't
5 unfairly benefit.

6 In considering this recommendation, note
7 that a single sequence prescribes only the
8 top-level curricular order for a grade. For
9 example, the topics that might be covered in
10 quarter 1 versus quarter 2. Within quarter
11 sequences and how to address the topics can be
12 left open. The alternative to a single sequence
13 is to require that each consortium submit a plan
14 for evaluating the comparability of scores from
15 different administration sequences, which could be
16 very difficult to meaningfully evaluate.

17 Question 2 concerns a system for
18 certifying the quality and rigor of end-of-course
19 assessments. The only recommendation I have for
20 that is that EDS bidders propose a method for
21 certifying that quality and rigor. That method
22 should include a proposed process and the evidence

1 to be used. For example, alignment of the
2 end-of-course test with common standards. A
3 comparative review of those tests against other
4 highly regarded end-of-course tests. For example,
5 international A levels. And a review of each end-
6 of-course test's technical characteristics.
7 Bidders should also be asked to propose a
8 qualified independent body to refine that process
9 to empanel experts and to conduct the reviews.

10 Question 3 concerns computer-based
11 assessment. I have three recommendations. The
12 first is that ED consider suggesting
13 computer-based assessment as a preferred model for
14 a significant component of the competition. That
15 preference might be based on the fact that
16 workplace and advanced academic settings routinely
17 require individuals to do cognitive work on
18 computer. And to the extent that common standards
19 reflect these requirements -- and I think it's
20 likely that they will -- paper testing might not
21 be able to measure the standards fully. However,
22 whatever the bidder's chosen model, ED should

1 require the bidder to justify the fit of that
2 model with common standards, as well as with the
3 other goals of the Race to the Top Assessment
4 Program.

5 Question 3 gives considerable attention
6 to comparability between paper and computer tests.
7 Why and when is comparability important?
8 Comparability is important when, for example,
9 assessment results are to be compared over time
10 and the delivery mode has changed from paper to
11 computer. If the scores are not comparable,
12 trends may no longer be interpretable.
13 Comparability is also important when assessment
14 results are to be compared across individuals.
15 And some individuals have taken the test on paper,
16 while others have taken it on the computer. If
17 the scores are not comparable, those comparisons
18 may be unfair.

19 Comparability is also important when
20 population groups are to be compared and the
21 proportions of students taking the test on
22 computer differ across those groups. If the

1 scores are not comparable, those group comparisons
2 may be meaningless. Note that if situations like
3 these don't apply, then there's no need to worry
4 about cross mode comparability. The dilemma
5 referred to in the text of question 3 is that
6 moving a large testing program to computers is
7 likely to require a multi-year transition. And if
8 maintaining cross mode comparability is important
9 over that period, innovation that threatens
10 comparability will be difficult to implement.

11 My recommendation is that ED require
12 bidders to propose a strategy for dealing with
13 that dilemma. And I'll offer two possibilities.
14 In the incremental innovation model, a consortium
15 would create parallel paper and computer tests
16 from the same content specifications. Collect
17 comparability data and equate the tests if
18 possible to render scores comparable across the
19 modes. It would then run those paper and computer
20 programs in tandem, transitioning more students in
21 schools to computer until paper administration
22 became the exception. At that point it would

1 introduce innovation that takes advantage of the
2 computer in ways that can't be duplicated on
3 paper. The advantage of this model is that it
4 preserves the meaning and fairness of score
5 interpretations that depend on comparability. The
6 disadvantage, of course, is that it delays
7 innovation until the out years.

8 In the concurrent innovation model, a
9 consortium would create innovative computer-based
10 assessments and obviously, noncomparable paper
11 tests. It would set performance standards
12 separately for each test. It would have a
13 representative sample take both tests and attempt
14 to create a concordance as exists for tests like
15 ACT and SAT, which might allow cross-test
16 comparisons and aggregations. It would run those
17 two programs, paper and computer, in tandem,
18 transitioning students to computer until paper
19 became the exception.

20 The advantage of this model, of course,
21 is that it advances innovation. The disadvantages
22 are that it may appear unfair to some as the paper

1 tests may seem inferior and may, in fact, be
2 inferior. And to the extent that the tests
3 measure considerably different constructs,
4 cross-test comparisons and aggregations may have
5 little meaning. That is, the concordance might
6 not work.

7 My last recommendation concerns the
8 evidence that should be supplied for comparability
9 in those cases where it's deemed important. Where
10 it's deemed important, the ED should require
11 bidders to provide evidence consistent with
12 professional standards. For example, the APA
13 guidelines for computer-based tests and
14 interpretations. In general, scores may be
15 considered equivalent when across modes the rank
16 orders closely approximate one another, implying
17 that the two modes are measuring the same
18 construct and when the score distributions are
19 approximately the same or have been made
20 approximately the same through statistical
21 adjustment, implying that the two constructs or
22 that construct is being measured on the same

1 scale.

2 Question 4 concerns how to encourage
3 ongoing innovation. My first recommendation is to
4 require that bidders present a long-term vision
5 for a next generation assessment system. That
6 long-term visions should include a rationale for
7 why that vision is meaningful, a set of steps to
8 progressively move toward it, a clear statement of
9 why the system developed under the Race to the Top
10 testing program would be a significant step toward
11 that vision, and a plan for continuing progress
12 toward the vision after the program funding ends.

13 My second recommendation is to require
14 that bidders present a specific plan for
15 continuous innovation during the funding period.
16 That continuous innovation plan should include one
17 or more existing assessment- or education-
18 innovation centers as consortium partners, and it
19 should closely involve students, teachers, and
20 administrators in design, tryout, and evaluation.

21 I would suggest considering the
22 following two continuous innovation models. The

1 first model is a school- based one. So select by
2 competition a subset of schools of varying
3 demographic characteristics to serve as assessment
4 innovation partners, designate them for a set
5 period -- three or four years, and give them a
6 waiver from accountability requirements that would
7 impede innovation.

8 The second model is a project-based one.
9 So select participating schools on a rolling,
10 project-by-project basis, which potentially allows
11 more schools to participate, but may create less
12 coherence and less critical mass. And of course,
13 the two models can be combined in a hybrid fashion
14 within a consortium.

15 Question 5 concerns focused research,
16 for which I have three recommendations. With
17 respect to value-added modeling, I'll read the
18 following from the Board on Testing and
19 Assessment's October 9, 2009, letter to Secretary
20 Duncan.

21 "A 2008 BOTA NAE Workshop on valuated
22 modeling concluded that there is little scientific

1 consensus about the many technical issues that
2 have been raised about value-added modeling
3 techniques and their use. BOTA agrees with other
4 experts who have urged the need for caution and
5 for further research prior to any large-scale,
6 high-stakes reliance on these approaches."

7 BOTA seems to be saying that ED fund
8 focused research. And I guess I would agree.

9 My second recommendation concerns
10 performance assessment. We know a lot about the
11 problems, but for use in accountability there's a
12 lot more we need to know about workable solutions.
13 For example, through-course assessment may be a
14 solution to the generalizability problem because
15 it distributes a large number of tasks and
16 including performances over time. But we need
17 research to confirm how well that idea would
18 actually work in practice.

19 As a second example, despite their
20 appeal we know surprising little about how to
21 create meaningful scores from unstandardized
22 projects and portfolios, especially if scored

1 locally. The Queensland Model and Singapore
2 Project work are often cited as examples, but
3 there is relatively little published research on
4 the technical quality of those assessment
5 programs.

6 Finally, we know very little about how
7 to score computer-based performance assessments,
8 which every mouse click, every keystroke, and
9 every resulting event are recordable. There's a
10 treasure trove of information about student
11 knowledge and skill in that trail, but we're only
12 just beginning to learn how to uncover it.

13 Finally, we know very little about how
14 to fairly assess students with disabilities and
15 those who are English language learners -- ELLs --
16 with the above methods. I would therefore
17 strongly recommend funding a program of focused
18 research in addition to main competition research
19 for consortia who want to use those approaches.

20 My last recommendation concerns learning
21 progressions. I believe, like Scott, that there's
22 considerable potential here for guiding assessment

1 design within and across grades and for providing
2 tentative formative feedback from summative tests.
3 But there are relatively few well-researched
4 progressions and no examples of their use in
5 design or reporting for large-scale assessment
6 that I know of. I would recommend a significant
7 focused research program to generate and
8 empirically support progressions in English,
9 language arts, and math, to incorporate them into
10 large-scale assessment design reporting; to create
11 related classroom assessments that can point
12 toward appropriate instructional materials; and to
13 evaluate the impact of the above.

14 In summary, I recommend for a
15 through-course summative assessment that ED be
16 suggested as a preferred model; require a theory
17 of action and associated research plan; suggest
18 that each consortium agree upon a single sequence
19 instead of administration windows or require
20 evidence to support the meaning of scores from
21 different sequences; and require a plan for
22 protecting the security of test content. For

1 certifying end-of-course assessments, I recommend
2 that ED require bidders to propose a certification
3 method, including a process and evidence to be
4 used and a qualified independent body to carry out
5 that process.

6 For computer-based assessment, I
7 recommend ED suggest it as a preferred model for
8 at least a significant proportion of the
9 competition and requirement is to justify the fit
10 of their chosen assessment mode, to propose a
11 strategy for dealing with the comparability
12 dilemma, and to provide comparability evidence
13 where needed that's consistent with professional
14 standards. For innovation and improvement over
15 time, I recommend that ED require bidders to
16 present the long-term vision, including a plan for
17 progressing toward it once the program funding
18 ends, and a specific plan for continuous
19 innovation during the program period.

20 Last, I'd recommend the funding focus
21 the research on valuated modeling; on performance
22 assessment, including aggregation methods; score

1 meaningful projects and portfolios; computer-based
2 performance assessment; and fairness for special
3 populations. And finally, on learning
4 progressions.

5 That's my input. Thanks very much.

6 MS. WEISS: With 27 seconds to go and an
7 incredible amount in the last 20 minutes, you two
8 have set very high bars. I'd be scared if I was
9 sitting at that side of the table right now.

10 So, questions? Go ahead, John.

11 MR. EASTON: Randy, I was really
12 intrigued by your first recommendation around the
13 through-course exams and the explicit explication
14 of the theory of action with a very strong
15 research and evaluation agenda built around it.

16 I'm just wondering if you could give
17 some examples -- points to some models for that
18 kind of work or examples for that kind of work.
19 My concern is around, you know, timeline issues
20 and feasibility of such an ambitious agenda.

21 MR. BENNETT: Well, you know, this is
22 going to be embarrassing because the first thing

1 that comes to mind is my own work. This is
2 essentially what, you know --

3 MS. WEISS: That's why we invited you.
4 It's okay.

5 MR. BENNETT: Oh, okay. We've been
6 working on through-course assessment -- or I
7 should say a type of through-course assessment
8 because there are different approaches -- for the
9 past four years. So, you know, we have been
10 developing an approach toward this that, you know,
11 we're trying to determine how effectively it could
12 be put into place. So that's one approach.

13 But there are other approaches to
14 through-course assessment, as many other have
15 noted in other countries. Queensland does a type
16 of through-course assessment that might also be
17 worth considering, but there's a lot less --
18 there's not as much -- there's not much research
19 about that either.

20 MS. WEISS: Was that your question? Did
21 that --

22 MR. EASTON: My question was looking for

1 a model for this intensive R&D model in the
2 development of the through course -- is basically
3 what you're recommending.

4 MS. WEISS: Other -- any questions? Go
5 ahead, Scott.

6 MR. MARION: Randy, just picking up on
7 John's point, I mean, it seems like this research
8 and evaluation effort of your theory of action --
9 I mean, some of us would call that sort of
10 validity argument. And it's really -- and there
11 are certain models out there. So, I mean, really
12 what you're doing -- you're posing claims and
13 testing those claims and looking to falsify them.
14 We are, I mean, we're pretty weak in that under
15 the NCLB assessments, although in the Special Ed
16 and the ELL assessments people have been doing
17 actually more work in that area because of
18 generosity from the Department in GSEGs and EAGs.
19 But really you're posing a validity argument,
20 right?

21 MR. BENNETT: Yeah. No, absolutely.
22 It's a validity argument with a lot of emphasis

1 put on the intended effects of the assessment
2 program as a mechanism for some sort of change in
3 either systems or students. If you read Michael
4 Kaine's chapter on validity in the most recent
5 edition of Education Measurement, he does include
6 that component in the validity argument. But it's
7 -- I won't say it's an afterthought, but it's
8 given secondary importance because most assessment
9 programs don't intend to affect change in
10 individuals and institutions. So, I've sort of,
11 like others, turned it around and given prominence
12 to that aspect of the assessment program. And if
13 you look at the theory around the program
14 evaluation, action -- theory of action takes -- is
15 given prominence there. So you're right. It's
16 just -- they're similar. Very similar.

17 MR. EASTON: It's also what Scott was
18 recommending on the value-added research -- on the
19 use of the VAM results. It's the same kind of
20 validity argument. Absolutely.

21 MS. WEISS: Go ahead, Laurie.

22 MR. WISE: One other source for a

1 potential model -- states have been required to
2 supply some sort of evidence about the
3 consequences of their assessment system. And I'm
4 not sure any state has done it really well, but
5 part of that -- you might look through what kinds
6 of evidence have been developed and supplied and
7 what kind of rationales have been stated that then
8 drive the research that brings in that kind of
9 evidence.

10 MS. WEISS: Other questions? So we got
11 one question from the audience that's a different
12 perspective than the ones we've been talking
13 about, which is whether there's a role for parents
14 in any of the proposed piloting of school-based
15 assessment models.

16 MR. BENNETT: Yeah, I think there
17 certainly is because parents are a consumer of
18 assessment results. And ideally you would want to
19 design a assessment program that would be capable
20 of giving parents assessment results that they
21 could, number one, understand; and number two,
22 have some possibility of doing something with.

1 So, I didn't -- and I thought about it -- I didn't
2 include parents. And I gave a list of the types
3 of actors that I thought a consortium should make
4 sure to work with in doing innovation. But
5 parents should be among them.

6 MS. DeSTEFANO: And I think the use of
7 theory of action provides a great role for parents
8 because when you involve parents you have to make
9 sure that you're putting them in a role where they
10 can be successful and they can be equal players.
11 And I think the use of theory of action is a good
12 niche for parents. And also, the usability
13 testing of reporting and performance descriptors
14 and other information that surrounds testing,
15 certainly involving parents in the testing of
16 that.

17 MS. WEISS: Thanks. Let's pass it on to
18 Jeff.

19 MR. NELLHAUS: Let's see how this works.

20 Well, good morning, everyone. It's nice
21 to be here. Thank you for inviting me. And it
22 also is a very -- two very tough acts to follow

1 here. They did a great job and I hope -- I'm
2 going to take probably a slightly different
3 perspective on some of the questions that were
4 asked. I actually tried to answer them as they
5 were stated. If this were required, what would
6 you expect applicants -- what would you ask of
7 applicants? So you'll see that my responses are
8 based on that sort of perspective.

9 My first -- I would like to make one
10 initial comment on the through-course assessment
11 system in that given the design of the system
12 where it will be component exams administered over
13 the course of the year and probably in some
14 particular sequence, I have to believe this is
15 going to have a significant impact on local
16 curriculum and instruction. Right now our state
17 assessment programs are agnostic about curriculum.
18 Curriculum -- the standards can be introduced at
19 any time in the year teachers and schools wish to
20 do so. This will force probably a much tighter
21 relationship between the standards, the
22 curriculum, and the assessments, which if that's

1 the theory here it may well do that. It may well
2 also help provide data more frequently to schools
3 to improve curriculum and instruction. So it does
4 have an upside.

5 The downside of the system as we see it
6 is that it may compete with what a lot of states
7 are now doing a lot of school systems with interim
8 informative assessment. So you're going to have
9 these three through-course high stakes assessments
10 competing with the low stakes interim and
11 formative assessments. So I'd just like to start
12 with that remark.

13 Now, my next few slides have to do with
14 the through-course question. And I tried to
15 address some of the questions on construct
16 validity, reliability, comparability, so on and so
17 forth. And the way I've structured these slides
18 is by creating an assumption or a premise on the
19 left. And given that premise or assumption, what
20 applicants should be asked to describe in their --
21 in response -- what you should ask applicants to
22 describe in the RFP.

1 So the first one has to do with
2 construct validity. And the construct that I want
3 to identify here -- and there's an assumption here
4 that proficiency will mean going beyond
5 demonstrating basic grasp of individual standards
6 of groups of closely related standards and include
7 the application of multiple standards from any
8 aspect of the content area to solve complex
9 problem. Now, if that's the construct -- if
10 that's the proficiency construct, I think
11 applicants should be asked to, one, reconfirm that
12 this, in fact, is their concept of proficiency and
13 then describe how their approach to its
14 measurement using the through-course assessment
15 system will accomplish measuring that particular
16 construct.

17 And I would suggest asking applicants to
18 address three questions here or assert that the
19 applicants should, one, state the standards that
20 will be assessed by each component exam; state how
21 each component will address standards assessed
22 previously so we're not just testing isolated

1 standards through the school year -- that you need
2 to go back and reassess some of the standards that
3 were addressed -- assessed previously. And then
4 thirdly, how individual test items will address
5 multiple, as well as single, standards. So we
6 don't get into a mode here, sort of a mastery
7 assessment mode where we're just trying to see if
8 students have mastered particular standards. The
9 construct is about how do we use all of these
10 standards together to solve complex problems.

11 So that's my comment about construct
12 validity. This is off the slide here, but my next
13 comment is on external validity. And the -- I
14 think an important measure of the external
15 validity will be the extent to which the
16 through-course assessment system will be able to
17 accurately report whether the students are on
18 track or college ready. You know, on track
19 meaning if they're in the 4th or 5th grade are
20 they ready for the next grade -- to take on
21 coursework in the content area of the next grade?

22 So if that's, again, the external

1 validity measure -- the most important one we're
2 looking at -- applicants should be asked to
3 describe, one, how they plan to report the results
4 of each component exam. Two, how they plan to
5 aggregate component results. As I think Scott
6 mentioned, the method they've chosen to aggregate
7 those results and why they think that's the best
8 method. And then thirdly, how they plan to
9 determine the summative score on each exam that
10 predicts readiness and how those scores will be
11 validated over time.

12 So I think these are just -- if the goal
13 here is to report readiness, I think these
14 measures need to be taken. I would also say that
15 they need to talk about in their plan their plan
16 for item development as well. Because in
17 measuring readiness the question will be are they
18 going to -- is the question whether or not the
19 student has met a standard? Or are we going to
20 try to report results across the whole performance
21 continuum? And that will have implications for
22 item development as well.

1 Whoops, I skipped one. Okay. Getting
2 back to the questions about reliability and
3 comparability. The assumption here is that the
4 level of reliability needed will depend on
5 reporting plans and intended uses of the results.
6 And that high levels of reliability will be
7 required for accountability uses. And thirdly,
8 that comparability requires high levels of
9 reliability and standardization of all elements of
10 the exam. Reliability is not enough to attain
11 comparability. So given those assumptions,
12 applicants should be asked to, one, describe how
13 they will achieve a level of reliability that
14 adequately supports the reporting plans and
15 planned uses of results.

16 So when it comes to reporting plans the
17 question is what kind of reports are going to be
18 provided for the component exams. Is it going to
19 be a raw score? Is it going to be a performance
20 level? Are there going to be subscales? So all
21 of those have to be requested of the applicants to
22 describe.

1 Secondly, the extent to which their
2 plans require standardized test administration
3 within and across schools and how that will be
4 achieved. Again, getting at the notion of
5 comparability is going to require standardization.
6 And if that's going to be a valued outcome of this
7 program, they need to describe the standardization
8 that will occur.

9 Thirdly, how they plan to establish high
10 levels of reliability and accuracy in the scoring
11 of constructive response questions within and
12 across years where they're scored externally by
13 contractors or locally by teachers.

14 And I know Scott mentioned that we have
15 a lot of methods to get reliability and accuracy
16 in scoring of constructive response questions.
17 Sure, when they're scored by contractors. But if
18 there's going to be teacher scoring here, that
19 creates an issue.

20 And then finally, their preliminary
21 plans for equating results across years. And
22 again, I would also say within years. So all of

1 these are factors that I think applicants should
2 be asked to describe to achieve reliability and
3 comparability.

4 One issue that wasn't addressed in the
5 questions that you gave us, but I think I have to
6 add something about it is feasibility. It's sort
7 of the third leg of a large- scale assessment
8 system of validity being one, reliability being
9 the other. But feasibility practicability is an
10 important feature of any large-scale assessment
11 program. And given that I think applicants should
12 be asked to provide an estimate of the average
13 yearly cost of the program, an estimate of the
14 testing time for each component, and a rationale
15 indicating that the testing time is sustainable.
16 An estimate of the LEA staffing time that will be
17 required to implement the program. And I would
18 also say an estimate of the SEA staffing time.
19 And finally, the amount of initial and ongoing
20 training and professional development that will be
21 required to launch and maintain the system over
22 time. These are all very important. Issues of

1 feasibility can't be overstated, if you will. I
2 think we can have very ambitious programs, but if
3 they're not practical, if they're not feasible,
4 they'll die on the vine.

5 In terms of the end-of-course high
6 school exams, the way I read the question was that
7 you are looking for a decentralized system. I
8 think Scott mentioned, you know, what's the common
9 about the common high school assessment program?
10 And in reading the question I assumed you were
11 looking at whether it was an individual school
12 district or even schools within the district would
13 try to establish end- of-course assessments, but
14 there would be some system that would certify the
15 rigorous quality of those assessments. And I
16 think maybe -- I believe the state of Pennsylvania
17 is actually trying to do something like that with
18 its high school exit testing program. So I'm sure
19 they've been thinking about this question quite
20 deeply.

21 But if that's the case, I believe
22 applicants should be asked to, one, describe the

1 criteria that will be used to certify the quality
2 and rigor of each exam in the set of end-
3 of-course tests that they're developing. But not
4 only of each exam, but across the set of exams.
5 Because you could easily have one test that's much
6 more rigorous than another if you don't look at
7 rigor across the set. And then finally, it gets
8 back to the construct validity question here.
9 What's the purpose of the end-of-course exams?
10 And if they're to predict readiness either for the
11 next most advanced high school course or readiness
12 for college or career, we need to -- the
13 applicants need to state the criteria that will be
14 used to valid that, in fact, those tests will be
15 able to measure readiness in that way.

16 Computer-based test administration. I'm
17 not an expert in this area, but, again, if it's
18 going to be required there are a number of issues
19 here. Implementation challenges, comparability
20 issues, and the testing of students with
21 disabilities. I'll just comment briefly on those.
22 In terms of implementation, applicants should be

1 asked to, one, describe how exams will be
2 administered in schools where computer-to-student
3 ratios are low and there is limited or no access
4 to broadband. This will be a big problem. And
5 also, their approach to ensure that students will
6 have the opportunity to learn computer-based
7 test-taking skills. So, how are they going to
8 ensure that students can keyboard? That they can
9 mouse? They can do the various things that will
10 be required when they take a test on the computer?

11 In terms of comparability with paper and
12 pencil exams, I'll just restate I think things
13 that were already mentioned. And that is that
14 applicants not be asked how they will ensure that
15 computer-based tests and any needed paper and
16 pencil versions assess comparable levels of
17 student knowledge and skill if preserving the full
18 power of computer- based test items is required.
19 Again, I think there's a competing issue going on
20 there where you don't want to discourage the
21 development of robust computer-based test items.

22 And then thirdly, as far as students

1 with disabilities goes, I think computer-based
2 assessments provide more advantages than
3 challenges for students with disabilities. So
4 applicants should be asked how they will take
5 advantage of computer-based assessments to improve
6 the validity of results for this population.

7 Just addressing the question on
8 innovation and improvement, I think there are a
9 number of things that applicants should be asked
10 to ensure that they're going to innovate over the
11 four-year period. First of all, they should be
12 asked to set aside a certain percentage of their
13 budget for research and development. There's not
14 going to be any research and development unless
15 this is a requirement. People will use the grants
16 they get for other things.

17 There be a requirement that the
18 applicants develop a four-year research and
19 development agenda identifying specific questions
20 applicants want to answer. Specifically,
21 questions that once answered would help them
22 innovate or improve the system.

1 Thirdly, they should be asked to
2 identify university and other partners who would
3 help them move their research agenda forward and
4 serve on advisory panels during the four years of
5 the grant period and beyond.

6 And finally, agree to share their
7 findings of the research with other states and
8 other consortia at periodic conferences sponsored
9 by the USDE and through USDE-sponsored research
10 networks. I'm sure there will be other venues,
11 but I think the Department of Education can play a
12 big role here to help support and promote
13 innovation and improvement.

14 And I just have one --

15 MS. WEISS: Go back one, I think.

16 MR. NELLHAUS: There we go. And issues
17 for focused research. Just a few quick comments
18 here. In terms of value-added research --
19 research on value-added methodology for teacher
20 and school accountability, this is an important
21 issue. It's a statistical modeling issue. Not
22 really a measurement issue. But I believe that

1 most states who have submitted their Race to the
2 Top applications have included doing some research
3 in this area. So the question is are you going to
4 have research being done independently from what
5 many states will be doing in terms -- in reference
6 to their own Race to the Top grant proposals? So
7 I would look to those proposals to see whether you
8 can build upon the research that's already
9 described in those proposals.

10 And then secondly, the second issue
11 there having to do with comparability,
12 generalizability in growth modeling for
13 assessments that include performance tasks. And
14 then I'll just restate what Randy and Scott have
15 already talked about in terms of, I think, the big
16 issue with performance tasks is equating --
17 including them in the assessments and creating
18 models for equating results that include
19 performance assessments over time.

20 So, with that I'll end my part of the
21 presentation.

22 MS. WEISS: Great. Thank you, Jeff.

1 That was terrific.

2 Questions?

3 MS. WURTZEL: Jeff, in your comments on
4 through- course assessments you raised the issue
5 that through-course assessments might compete with
6 interim assessments that states and districts are
7 using now. And I just wanted to press on that a
8 little bit. We heard at some of the prior
9 meetings suggestions that the components of a
10 through-course assessment might replace interim
11 assessments or model what higher quality interim
12 assessments could look like. Do you see those as
13 both possibilities or do you feel like they're
14 very much in conflict with each other?

15 MR. NELLHAUS: Well, if the theory is
16 that people will have -- will frequent data from
17 the component assessments and use it like they do
18 interim assessment data, I think that's a theory
19 that needs to be tested because I think there's a
20 difference in using data from high stakes tests
21 and using data from low stakes tests. So, I'm not
22 sure whether the through-course assessment system

1 can replace the role that interim assessment
2 programs can play.

3 I will also say -- my other fear is that
4 you already -- one of the big criticisms of
5 large-scale assessments for accountability is that
6 it's narrowing the curriculum. You know, that
7 there's too much, you know, other things about
8 school are not being seen as important. And now
9 we're going to have that happening four or five
10 times a year in a school. And it could exacerbate
11 that issue of narrowing the curriculum. With an
12 interim assessment with low stakes it may mitigate
13 that tendency to try to just teach to the test.

14 MS. WHALEN: Can I ask just two quick
15 clarifying questions? When you were talking about
16 the end-of-course high school exams you mentioned
17 set of exams. Did you mean a set of exams within
18 an individual course or a set of exams across the
19 subject area?

20 MR. NELLHAUS: I meant a set of exams,
21 you know, different exams in different subject
22 areas.

1 MS. WHALEN: So in science you would say
2 chemistry, biology, physics -- the rigor across
3 all of them and how they all fit together?

4 MR. NELLHAUS: Right. Or, you know, how
5 does the rigor of the English test compare to the
6 math test?

7 MS. WHALEN: So across subject areas as
8 well.

9 MR. NELLHAUS: Across subject areas.
10 And again, it has to do with how you're going to
11 look at those sets of exams. Are you going to
12 certify or somehow indicate a student is college
13 ready or career ready with one test or with a
14 series of tests? So how are you going to use
15 groups of tests together to make those
16 determinations?

17 MS. WHALEN: And then my other
18 clarifying question was when you were talking
19 about the reporting of -- the reporting plans of
20 data. You mentioned (inaudible) scores,
21 performance level scale scores. Can you talk a
22 little bit about your thinking behind why they

1 would individually make a difference and what are
2 the advantages and disadvantages associated with
3 thinking about the levels of reliability with
4 each? Or did you mean something different?

5 MR. NELLHAUS: Well, if you think of
6 these through- course assessments, are they going
7 to take one hour? Two hours? I mean, right now
8 most state assessments take, in any given content
9 area, about two hours. They're usually two or
10 three sessions. They take an hour and a half to
11 two and a half hours long to administer. Now
12 you're going to be putting together component
13 assessments and it's unclear to me just what the
14 purpose of each of them will be. Will it be to
15 actually report out the extent to which students
16 are proficient on a particular subset of
17 standards? Or will it be just to -- maybe the
18 test won't be robust enough to do that. Maybe
19 they'll be looking at just -- we're just looking
20 at rolling up the results over time -- over the
21 course of the year to make some kind of summative
22 proficiency determination. So all of that needs

1 to be thought through by the applicants ahead of
2 time so they can talk about just, you know, how
3 comprehensive each of the components will be.

4 MS. WEISS: So, Jeff, you started with
5 the sort of don't forget the consequences of doing
6 this through-course thing that, you know, one
7 consequence might be forcing a tight relationship
8 between standards curriculum and instruction.
9 Another consequence might be losing or sort of
10 crowding out low stakes interim assessments that
11 might be important for instructional improvement
12 purposes. And then you went on to say so. Given
13 that if you really required them here is what I
14 think, so take a step back and tell us whether you
15 would require them or not, like how you balance
16 these tradeoffs given what you know about the kind
17 of instructional improvement you're hoping to
18 enable at the school and classroom level.

19 MR. NELLHAUS: Yeah, well, you know,
20 thinking about doing this in grades 3 through 8
21 just seems overwhelming for a school, so I'll
22 start there. I just think it's very, very

1 ambitious. So if we could start out in a
2 particular subject area or a particular grade
3 level to test this out I would be much more
4 comfortable. Also, I like the idea of having some
5 curriculum-embedded tasks and exercises eventually
6 rolled up into a summative score. I think that's
7 a great idea. But I think we need to be strategic
8 and parsimonious about how we do it.

9 So, for example, there are certain
10 things that just don't lend themselves well to
11 on-demand testing at the end of the year. You
12 know, conducting a scientific investigation,
13 writing a research paper, giving an oral
14 presentation. These are things that we don't do
15 well in on-demand assessments. Those could be
16 incorporated at strategic times in particular
17 grade levels and include that in summative
18 results. I think that would be a great step.
19 That would be a step in the right direction in
20 terms of trying to get more information into the
21 summative score. But I think the idea of doing
22 through-course -- I would say it might be -- I'm

1 not -- I'm trying to understand -- think about
2 where it might lend itself more useful. And I
3 know programs that have this -- like the
4 International Baccalaureate, the Cambridge
5 International -- I know these are models for doing
6 this sort of thing. They tend to be done at the
7 middle or especially at the high school level.

8 You know, how well this is going to work
9 at the elementary and middle school level I think
10 is a big question right now. So I like the idea.
11 It could be an option. You know, I think we have
12 to look behind what's the theory of action here.
13 Is this the best means to accomplishing whatever
14 the goal is that we have? And if the goal is
15 simply just to have more frequent data during the
16 school year for which to improve curriculum
17 instruction, you have to ask yourself whether this
18 is the best way to accomplish that.

19 And in terms of feasibility, what is the
20 cost of this going to be? I know most states now
21 -- and maybe we'll get out of the economy we're in
22 eventually, but I can't imagine this is going to

1 cost any less than what we're doing right now.
2 And it'll probably cost a lot more.

3 MS. WEISS: Thanks. Any other
4 questions?

5 MR. MARION: I think it's worth
6 emphasizing it's not just the development. It's
7 the ongoing cost that'll cost a lot more. And
8 that's a worry of the states.

9 MR. NELLHAUS: Right.

10 MR. BENNETT: The theory of action also
11 should identify potential unintended consequences
12 and how the consortium will attempt to forestall
13 those unintended consequences from occurring.

14 So, for example, it's very important to
15 be concerned about the potential for narrowing the
16 curriculum. And one might propose in one's theory
17 of action that those periodic assessments or
18 through-course assessments be designed to model
19 good instruction, encourage teachers to attend
20 more broadly to standards than they might
21 otherwise be doing, and so on. So, there really
22 ought to be an explicit attempt to identify in

1 advance what some of those concerns would be and
2 make sure to design into the assessment system
3 ways to prevent those from occurring.

4 MR. WEISS: Yeah, Jamal.

5 MR. ABEDI: (inaudible) in addition to
6 many other important issues, issues concerning
7 motivation. And many of those components of the
8 through-course are lower stakes-type of
9 components. And they don't have the higher stakes
10 attached to their state assessment that currently
11 we have. So that's an issue as well and it is
12 very important to pay attention.

13 MS. WEISS: Great. Thanks. Should we
14 pass it on to Lizanne?

15 MS. DeSTEFANO: Well, good morning,
16 everyone. And thanks for this opportunity.

17 Let me say with regard -- oops. There
18 we go -- with regard to the through-course
19 assessment issue, I believe that a through-course
20 assessment system, if it's well designed and well
21 implemented has many advantages over a traditional
22 end-of-course or end-of-year assessment system.

1 Multiple opportunities for assessment, varied
2 types of assessment, do much to increase -- can do
3 much to increase the reliability, the validity,
4 the fairness, the accessibility of an assessment
5 system.

6 However, as my colleagues have
7 indicated, there is a lot that we don't know about
8 through-course assessment systems. And poorly
9 designed or poorly implemented systems I think can
10 have a number of negative consequences that we
11 want to guard against. So with regard to the
12 requirement to implement a through-course
13 assessment system, I think that my preference
14 would be to offer it as an option to encourage
15 states that are interested in doing something to
16 pursue that, but to not have it be required
17 because I think there are a number of other models
18 that we would like to encourage and incorporate
19 that can come out with the same ends. And
20 through-course is not the only option, I think, to
21 improve our assessment systems.

22 One of the biggest issues I think in

1 considering all of the Race to the Top assessment
2 initiatives is to really keep grade level or
3 college and career readiness standards at the
4 center of the system, rather than letting the
5 assessments drive the system. The message should
6 be that we are not preparing students to pass the
7 tests, but that we are preparing students to gain
8 the knowledge and skills that they will need to be
9 successful in a particular content area.

10 And so when I think about advice to
11 states or reviewing the Race to the Top assessment
12 proposals, my inclination would be to look for
13 ways in which states and consortia of states are
14 keeping standards at the center. Not to take
15 standards as a given, but to actually spend a part
16 of the grant looking at their standards and, in
17 particular, their performance descriptors and
18 making sure that these are really -- embrace the
19 vision that they have for students in their state
20 and very much are externally valid connections
21 between the various levels of the system: The
22 readiness for kindergarten, movement from

1 elementary school to middle school, transition to
2 high school, and transition to college or career
3 readiness.

4 An importance piece of this, I think, is
5 involvement of higher education representatives
6 and business leaders in the examination of
7 standards and the development of performance
8 descriptors. I focus on that transition into work
9 or the transition into postsecondary because I
10 think it's one of the weakest links in our
11 standards and accountability systems. In my
12 experience, higher education and business have
13 been involved in standard setting and development
14 of performance descriptors, but in rather
15 superficial ways. Higher education
16 representatives are often brought in as content
17 experts for a particular area, but not really
18 looking at the articulation or the transition of
19 students from the secondary system into their
20 environment.

21 Likewise, business representatives that
22 have been involved in my experience are often CEOs

1 or businessmen or business people who are strong
2 advocates for education and not really the HR
3 people or the people who really have a deep
4 understanding of the skills that are necessary to
5 be successful in a particular position. And so I
6 would like to see states and consortia of states
7 think about how can we get standards and
8 performance descriptors that allow us to move
9 forward with this big vision. That's why a theory
10 of action is very appealing because those
11 standards and performance descriptors can underlie
12 that theory of action.

13 With respect to the through-course
14 summative assessments then, I think clear
15 specification of the constructs or skills that are
16 important in a particular course or in a
17 particular year of study and then linking those
18 assessments -- the through-course and the
19 summative assessments -- to those standards and
20 skills. The content validity piece is very
21 important. It's not trite. It's not pro forma.
22 It's not something everybody does, but really

1 thought needs to go into that.

2 And in addition to specifying the target
3 skills, okay, what are we really interested in
4 assessment specification of the access skills?
5 What skills do students need to be able to get to
6 those target skills are very important for
7 teachers so that they understand what they need to
8 scaffold or support their students to do in
9 classes. It's very important for the articulation
10 across grade levels and across educational --
11 different education institutions.

12 But also it's very important for
13 including English language learners, students with
14 disabilities in assessment and supporting them to
15 really demonstrate what they know about the target
16 skills and to try to accommodate or mediate the
17 deficits that are caused by access skills. So,
18 for me an important piece of construct validity is
19 not only focusing on target skills, but what
20 access skills do you need to be able to assess
21 that.

22 And then the same sort of accessibility

1 considerations that we think about in high stakes
2 assessment, end-of-course, end-of-year assessments
3 need to follow through with the through-course
4 assessments. Are we designing through-course
5 assessments in ways that allow a broad range of
6 students to participate in them and to be
7 successful? So I think those would be things that
8 I'd like to know that a state is taking care of.

9 The other thing that I think is very
10 important is to have a state or consortia
11 acknowledge the types of professional development
12 that are going to have to go on at all levels of
13 the system to enable teachers and principals and
14 parents and other educational professionals to be
15 able to make maximum use out of a through-course
16 assessment. And in my opinion, if a
17 through-course assessment because of security
18 issues or standardization issues -- if the
19 instructional utility and the real-time use of
20 through-course assessments is limited because
21 they're included in a summative assessment or in
22 an accountability system, then for me that would

1 diminish the utility of a through-course
2 assessment tremendously. Many states and
3 localities have made big gains in implementing
4 formative assessment that gives teachers
5 information that they need to make instructional
6 decisions. I agree with Jeff that the
7 introduction of a through-course system could
8 compete with that system in teachers' minds in
9 terms of time in the classroom, in terms of many
10 things. And so if there's not an efficiency and
11 if those interim assessments cannot also be used
12 for instructional purposes, then to me that would
13 be a deal breaker with through-course.

14 When we're thinking about a
15 through-course assessment, one of the advantages
16 particularly to ensure precision in the full range
17 of the performance continuum is to design the
18 system with that idea of maintaining high
19 precision across the full performance continuum.
20 And as Randy and others have said that that be an
21 explicit goal of this system so that we realize
22 that that end-of-the-course one-shot assessment

1 cannot adequately measure the full range of
2 performance. And we put in the interim
3 assessments as a way of bolstering precision at
4 different parts of the performance continuum,
5 primarily at the low end and the high end where
6 our instruments tend to measure quite poorly.

7 So modular or adaptive assessments that
8 allow for some differentiation within the
9 classroom. Different students might do different
10 things on those interim assessments to allow you
11 to really get the full range of performance. To
12 me that would be an elegant design.

13 And then finally, we don't know much
14 about the predictive validity of these
15 assessments. We don't know very much about the
16 consequential validity of these assessments and so
17 a lot of attention to those, and particularly the
18 aggregation rules. That's really the unknown area
19 here. We don't have a lot of experience. And so
20 a strand of research that maybe floats above the
21 consortia so that the consortia are evolved, but
22 there's a meta entity that's looking at these

1 issues across the consortia might be one way to
2 deal with that.

3 The second issue -- system for
4 certifying quality and rigor end-of-course exams
5 -- once again, the centrality of the standards and
6 the performance descriptors are really important
7 here. To me, if you have good standards and
8 performance descriptors and you know what is to be
9 taught in each course, it goes a long way to
10 validity. It goes a long way for comparability.
11 It goes a long way with standardization. And so I
12 think, again, the upfront work to me is very, very
13 important. And here predictive and concurrent
14 validity are essential. So looking at the extent
15 to which these end-of-course exams predict success
16 in future courses, in work, in postsecondary for
17 varying groups of students. So not a kind of one
18 size fits all, but to identify target groups and
19 take a look at the validity of these exams for
20 predicting future success. Concurrent validity,
21 of course, with other commonly accepted end-of-
22 course exams is really important.

1 This is a place where higher education
2 partners can be very, very useful. At my
3 university we've had a lot of dissatisfaction with
4 AP exams, particularly in math. We don't feel
5 that they predict student success very well in the
6 calculus at our university. And now we're working
7 in partnership with the State Board of Education
8 and some school districts to look at the
9 relationship between curricula in those school
10 districts; performance on local, state and AP
11 exams; and success in calculus at the university
12 level. Very, very important and very reciprocal.
13 It's helping the university faculty and the
14 university leaders to understand more about what's
15 going on in community colleges and high schools.
16 It's also helping community college and high
17 school teachers to understand the expectations of
18 the university. And so it's been a very fruitful
19 process, I think, for helping with articulation
20 issues.

21 Computer-based test administration.
22 Again, I think this is the future. I think I

1 would be surprised if states did not have some
2 aspect of computer-based or computer- adaptive
3 testing in their proposal. But again, requiring
4 it -- we may not be at the point where we're
5 talking about requiring it at all grade levels in
6 all subject areas. So, to me, a more
7 well-reasoned proposal would be a proposal that
8 picked a particular area that the state really
9 wanted to invest and learn about computer-based or
10 computer-adaptive testing and then take what they
11 learned there and roll it out to larger numbers of
12 schools, larger numbers of students, larger
13 numbers of grade levels or performance areas.

14 A couple of particular things that are
15 about computer-based testing or computer-adaptive
16 testing is that -- particularly for subgroups are
17 that novel item types may differentially
18 disadvantage students with disabilities and other
19 subgroups. Some of the work that my colleagues
20 and I have done with NAEP found that if students
21 with disabilities are encountering item types for
22 the very first time, it may be very difficult for

1 them to discern what's important, how to approach
2 that item. And also it increases the amount of
3 time and the anxiety associated with testing a
4 great deal.

5 So I think we have to -- in embracing
6 novel item types and recognizing that computers
7 can help us get lots of wonderful information
8 about what students know and can do, that we also
9 have to make sure that there's a strong link
10 between the kinds of things that students are
11 exposed to in instruction and the kinds of things
12 that they see in assessment. And that this may --
13 what may be true for students who take -- put
14 their courses predominantly in mainstream classes,
15 may not hold true for students in special
16 programs, like English language -- programs for
17 English language learners or for students with
18 disability. So that link between what they
19 encounter in instruction and assessment is very,
20 very important.

21 Likewise, for teachers. I put students
22 should have ample experience with technology used

1 in assessment, but I just went to a training. A
2 district that I work with is rolling out a
3 computer-based assessment system in grades 2
4 through 8. And I went to a teacher workshop last
5 week and I was struck -- I think I know this
6 district very well, but I was struck by the range
7 of computer skills that the teachers had in this
8 training. And these were volunteer teachers.
9 These are teachers who want to be in the first
10 phase of implementation of this assessment. But
11 there was still a very wide range of computer
12 knowledge among this group of teachers.

13 So, again, in addition to thinking about
14 what should the test look like, also states and
15 consortia should be thinking about what
16 professional development and support,
17 infrastructure and training, needs to be provided
18 for these to be successful. Not to say that this
19 is a negative in terms of developing
20 computer-based assessments, but to say that this
21 is a realistic problem that we have to address.

22 And then finally, computer-adaptive

1 testing algorithms that are developed on
2 mainstream populations may not be appropriate for
3 students with disabilities or English language
4 learners who tend to have more splinter skills
5 rather than linear learning progressions. And so
6 the research on learning progressions really needs
7 to include attention to subgroups when designing
8 computer adaptive tests.

9 With regard to innovation and
10 improvement beyond the four-year grant period, as
11 everyone in this room knows, a huge barrier to the
12 effective implementation of No Child Left Behind
13 was the lack of time for R&D. You know, all of
14 the time we were building the plane while we were
15 flying it and trying to figure out how to meet the
16 letter of the law dealing with many other
17 competing challenges that we had. So, it would be
18 a huge advancement if this competition should
19 really take that into consideration. And the
20 Department should strongly encourage applicants to
21 be realistic about what can happen in the
22 four-year period of time. And the criteria and

1 review process should value a realistic approach.

2 So, Scott started out with Lorrie
3 Shepard's idea about targeting areas, targeting
4 grade levels, picking small things that you can do
5 well. And I would encourage that the Department
6 and the reviewers and the states take that very
7 seriously with the idea that there's plans for
8 scaling up. There's plans for going to scale.
9 There's plans for moving into other content areas,
10 other grade levels, other assessments.

11 So, I would specifically encourage that
12 applicants -- given this theory of action, this
13 big vision -- submit a very detailed management
14 plan for the four years of funding that lays out
15 R&D; the products that will be developed during
16 that period of time; what will be implemented; and
17 how this information will be used within the four
18 years. But they also be required to include an
19 out years plan. It may be less detailed, but it
20 includes then, okay, once federal funding is over
21 how will these activities include -- be carried
22 out? How will they be funded? And what's the

1 ultimate goal of this assessment system?

2 And finally, for additional focused
3 research, I agree that considerable research
4 exists on growth modeling and value-added
5 methodology, although again I can see a value for
6 some sort of a meta structure that supports the
7 work that's going on within the individual
8 consortia. Something that raises the findings and
9 the successful practices of the consortia to a
10 level that other consortia can benefit from it.
11 And also to provide guidance and support to people
12 within states and within the consortia to keep the
13 work going.

14 Sometimes State Departments of Education
15 and other partners are not the best prepared to
16 have these ongoing systems of R&D because they're
17 always getting subverted by the practical, okay,
18 we've got to test on this particular day. We've
19 got to get the report out here. And so having a
20 meta structure that helps stabilize that and keep
21 that work moving forward I think would be very
22 important.

1 I think studies of use or nonuse and
2 consequential validity of the various types of
3 higher mid-stakes assessment are needed. I think
4 we all have presumptions of use, but I think we
5 might be surprised if we actually got down into
6 classrooms and into schools and into districts to
7 see the types of decisions that are made. I think
8 that would be an important line of inquiry.

9 And then finally, the studies of the
10 functioning of the consortia. Consortia are a
11 requirement in this competition. Some of us have
12 participated in consortia, and I think we all
13 would agree they have various degrees of success.
14 Some are highly functional; some are
15 dysfunctional. So I think a line of research that
16 kind of looks at the consortia functioning and
17 again tries to feed back best practices to
18 consortia to keep them fresh. Four years is a
19 long time to collaborate, so some mechanism for
20 keeping things fresh, keeping things moving,
21 allowing for change I think would be very
22 important.

1 Thank you.

2 MS. WEISS: Thanks, Lizanne. Questions?

3 MS. WURTZEL: So, Lizanne, you raised
4 the issue that adaptive testing may be problematic
5 for students with disabilities because of their
6 different patterns of knowledge. So if a
7 consortia were to propose to do adaptive testing
8 what kind of research or validity work would you
9 want them to come in with or propose to do over
10 time to make sure that their adaptive testing was
11 appropriate for that population?

12 MS. DeSTEFANO: I'm actually glad that
13 you asked me that question because I didn't mean
14 to come out negative with computer-adaptive
15 testing or computer-based testing for students
16 with disabilities because I actually believe
17 there's a lot of value and a lot that can be done
18 to increase accessibility. So I realize now my
19 comments might have been a little bit balanced.

20 For example, I think a lot of people
21 leap into the computer-adaptive testing boat very
22 quickly. I think there's a lot of negative issues

1 there for students with disabilities. I'll talk
2 about that in a minute.

3 On the other hand, I think the computer
4 can offer a lot of opportunities to increase
5 accessibility. I think the read-aloud features
6 for students with disabilities are wonderful with
7 computers. And I was very disappointed in this
8 school district that I told you about that is
9 implementing computer-based assessment, that they
10 don't have a read-aloud option. There's no voice
11 at all in the computer administration. I think
12 that can help a lot with students with reading
13 problems, students with attention issues, to have
14 an audio component of an assessment. It can help
15 pace them. It can help them with words they don't
16 know. A thesaurus or other kind of function can
17 be very, very useful.

18 So I think that -- also information
19 about errors that can be gained from
20 computer-adaptive testing can be very useful for
21 teachers working with certain groups of students
22 who have educational challenges. It can give them

1 very detailed information that you can't get from
2 a paper and pencil. So I didn't mean to be
3 negative. And I would really encourage developers
4 to fully explore the adaptive functions of
5 computers in terms of size of print, one item per
6 page, voice, other kinds of things that can really
7 help interpretation.

8 In terms of the negative consequences, I
9 think that cognitive labs -- a program of
10 cognitive lab research that includes
11 overrepresentation of English language learners
12 and students with disabilities in looking at
13 computer-adaptive testing is very, very important,
14 both from how the student interacts with the
15 computer on the day of testing, but also in
16 looking at the learning progressions and how well
17 the computer test compares with other sources of
18 information about that student's performance. Are
19 we doing a good job in computer-adaptive testing
20 in capturing the learning progressions of students
21 with disabilities? And I would put English
22 language learners or students of poverty in that

1 category, too. Or are we somehow misrepresenting
2 or underrepresenting what they know and can do?

3 A lot of the tests that I've seen in
4 computer- adaptive testing use vocabulary-types of
5 items to kind of place students in the performance
6 continuum, particularly in literacy. And again, I
7 haven't done research in that area, but I think
8 for some types of students vocabulary is not the
9 right skill to get a very general sense of where
10 the student belongs in the continuum. We should
11 be looking at other types of item.

12 MS. WEISS: Other questions? Hearing
13 none and knowing it's lunchtime, we will take a
14 break. And why don't we see you back in here at
15 -- we'll reconvene --

16 (Whereupon, at 11:59 a.m., a
17 luncheon recess was taken.)

18

19

20

21

22

1 with the following characteristics: A, a system
2 that includes components of assessment delivered
3 periodically throughout the school year. If we do
4 this, how should we ask applicants to describe
5 their approaches? What evidence should we request
6 if such summative results are part of the
7 accountability system?

8 So I start with the first question,
9 Section A. And I wanted to start with the
10 potentials of the through-course assessment
11 provides value-added components for a more
12 thorough assessment of a student's performance;
13 provides multiple measures from different sources
14 with different formats, performance assessment and
15 so forth. Accountability system will be based on
16 a more comprehensive picture of what the students
17 know and are able to do, and provides bigger
18 assessment opportunities for special needs
19 students, especially for English language learners
20 because one of the major issues for English
21 language learners is having one opportunity to
22 present what they know and can do, which is really

1 a little and they need more opportunity to
2 express. And this gives them more opportunities.
3 So I'm really happy about this, especially
4 teachers getting involved and more opportunities
5 for more components.

6 However, all these good things have some
7 consequences and there are some challenges. The
8 first challenge is extra effort and expense in
9 developing and implementing the new components,
10 extra time taken from instruction, comparability
11 (inaudible). Anyways --

12 MS. WEISS: Maybe hold it up.

13 SPEAKER: It's very sensitive.

14 MR. ABEDI: Okay, great. Comparability
15 issues in content construct, linguistics, and
16 psychometrics of the different components within
17 and across the consortia of the states; burden on
18 teachers and the schools or whoever is responsible
19 for providing data; and even more importantly,
20 issue of redundancy of information. It is
21 important to understand and see and we have to be
22 aware that we don't want to include many

1 components that are all measuring the same thing.
2 So we need to know of the specific contributions
3 of each of those components that are going to
4 make. Because if they are measuring the same
5 components, same thing, same construct or same
6 content, then what is the purpose of including
7 more than one measure? So the issue of redundancy
8 is an extremely important issue to discuss here
9 and to see the different components actually have
10 a contribution (inaudible). Second question would
11 be how should we ask applicants to describe their
12 approaches. I am assuming that the consortia
13 actually are going to be responsible for defining
14 these components. Clearly, identify the through-
15 course components and the relevance to the
16 assessment and accountability system. What are
17 those components? Define them. Clearly document
18 the value added by the through-course components
19 to the assessment and accountability system.

20 These (inaudible) field test methodology
21 for creating and interpreting the composite
22 systems. My colleagues talked about how to create

1 a composite of those components. As Scott
2 mentioned, for instance, it may not be a good idea
3 to add them together. Just a simple composite.
4 However, depending on the level of their
5 contribution it is important to find some ways to
6 actually combine them and come up with some kind
7 of general, oral components.

8 Identify burden on teacher, school
9 districts, consortia of states, or whoever is
10 responsible for collecting the components and
11 incorporating them into assessments. And then
12 present evidence on the problem of redundancy on
13 added criteria. Again, there are two important
14 issues. One of them, how much unique contribution
15 each component would have and then if there are
16 any kind of redundancy. Again, states and schools
17 have enough tests to do right now. We don't want
18 to add additional tests unless we know that they
19 are going to do good or they are going to
20 contribute and they are going to help us with the
21 new concept and adding to the -- improving the
22 validity of the assessments.

1 The second part, okay, see, what
2 evidence should we request if such summative
3 results are part of the accountability system?
4 Talking about validity, external validity for
5 postsecondary preparedness, reliability, fairness,
6 precision, and comparability. And I'm going to
7 elaborate on each of those things very, very
8 briefly.

9 First of all, on the Section A, what
10 evidence should we request for validity, including
11 construct validity, content, consequential, and
12 predictive validity? For construct, different
13 components measure -- again, this is an issue of
14 comparability and there are some major issues with
15 it. Measure the same construct. Construct
16 irrelevant sources of control. The components
17 would provide added values. The issue of
18 construct irrelevant variance or sources is
19 extremely important for students with disabilities
20 and English language learners because there are,
21 for instance, one of the major issue -- one of the
22 major construct irrelevant source for -- is

1 through English language learners -- is the
2 unnecessary linguistic complexity of assessments.
3 Assessments that are complex in language may not
4 give them a good opportunity to express what they
5 are unable to do. Because if they don't
6 understand the language of assessment, they may
7 not be able to provide their content knowledge to
8 express their content knowledge. So that's a good
9 example of construct irrelevant.

10 So it is extremely important to identify
11 -- for the consortia to identify these sources and
12 provide advice on ways for controlling these
13 sources. For content, evidence of alignment of
14 components, end-of-course and the through- course,
15 within the common set of K through 12,
16 internationally benchmarked, college and career
17 readiness -- ready standards.

18 Issues concerning consequential
19 validity, evidence on achieving the goal of
20 measuring a common set of standards examining both
21 intended -- for instance, what are the intended
22 consequences, determination of (inaudible) college

1 and career readiness, high school exit decision,
2 college course placement, and especially so
3 important to make sure of unintended consequences,
4 for instance, focusing on limited outcomes, the
5 possibilities of teaching to the test. There are
6 many other sources of unintended consequences have
7 to be. And I think the consortia that the
8 application should be clear on both intended and
9 unintended consequences. And also regarding the
10 predictability (inaudible) in predicting college
11 and career readiness.

12 On Section B, what evidence should we
13 request, external validity for postsecondary, I
14 wanted to go through this quickly. Suggestions,
15 measures, or indices of college and career
16 readiness, high school completion, college
17 performance, maintaining high performance at
18 college, and the students' attributes for college
19 and career readiness, these are some of the
20 components -- some of the criteria for predictive
21 validity.

22 Going to what evidence should we request

1 for reliability, including inter-rate reliability.
2 I divided this section into two different
3 sections. One of them is related to reliability
4 and one of them related to inter-rate reliability
5 because I believe the use of inter-rate
6 reliability is extremely important, especially
7 having more attention to open-ended -- extended
8 open-ended questions. So I wanted to address that
9 in a little bit more detail.

10 So the issue of reliability, identify
11 sources of systematic error of measurement or
12 bias. Again, maybe I, myself, am biased because
13 my focus of research has been on the students,
14 English language learners, and there are many
15 sources of bias for these kids, including cultural
16 and linguistic biases. So that's why I'm saying
17 that identify sources of bias or sources of
18 systematic error for measurement is important in
19 the reliability.

20 Test reliability, internal consistency,
21 examine dimensionality. One of the, you know, the
22 internal consistency approach is one of the most

1 used and misused approaches in estimating
2 reliability because one of the major requirements
3 and major issue in internal consistency approach
4 is dimensionality. It is extremely, as you all
5 know, internal consistency approach is extremely
6 sensitive to dimensionality. If the subscales --
7 if there are subscales that are not highly
8 correlated, then internal consistency is going to
9 underestimate reliability. But this is one of the
10 most commonly used approach. So what I'm trying
11 to suggest in addition to the internal consistency
12 approach, other approaches such as tests
13 (inaudible) reliability, parallel forms and stuff
14 like that. So consortias (sic) should plan for
15 additional approach, not solely relying on the
16 internal consistency approach.

17 Evidence -- this is extremely important
18 that evidence for reliability for subgroups --
19 there are many studies have done nationally and
20 shown that many of the existing tests are
21 extremely good, very high reliability, but when
22 reliability is computed for the mainstream

1 students. But when you go to subgroups of
2 students, then reliabilities are extremely low.
3 In some cases we found reliability of 0.5, 0.55
4 for math, reading, language arts, and so forth.
5 So when a test has a reliability of 0.5, it's just
6 like flipping a coin, a lot of measurement error
7 involved. So it is important that the consortia
8 actually have a plan to report and compute
9 reliabilities by subgroups as students with
10 disabilities, English language learners. Even
11 within students with disabilities, I don't know if
12 Lizanne would agree with me that different
13 categories actually perform quite differently. So
14 combining them, aggregating them into one single
15 group might not really work at all. So reporting
16 by subgroups -- within subgroups that are
17 different in terms of their background
18 characteristics.

19 Even ELL. When we are talking about
20 ELL, we are thinking that we are talking about
21 these very homogeneous groups, which is absolutely
22 not the case. ELL students are different in every

1 imaginable aspect, including the level of English
2 proficiency. Some of them you may not even
3 believe have higher level of proficiency than
4 native English speakers (inaudible) English
5 language proficiency -- I mean, ELLs. We don't
6 know, our -- the classification system doesn't
7 look. So that's a major national issue that when
8 consortia are going to talk, should address those
9 issues, what they do, whether they -- what they do
10 with assessment of subgroups with different
11 background characteristics, whether they are going
12 to -- their assessments are going to be sensitive
13 to those background stuff. And also, a standard
14 error of measurement should be reported.

15 Going to inter-rate reliability, as I
16 mentioned to you, I wanted to discuss that in just
17 another -- as another topic or title. What
18 evidence. Clearly identify factors affecting
19 validity and consistency of scoring open-ended
20 test items such as raters' background and raters'
21 experience. Involve teachers in the scoring of
22 open-ended items, but at the same time estimate

1 and discuss burden on teachers in getting involved
2 in ratings.

3 It is important because if you want to
4 ask teachers to get involved, they should receive
5 some training, obviously. They should know
6 exactly what they are doing. So all of this is
7 going to take time from their instruction. And so
8 how do you want to deal with that even though it
9 is extremely important to involve teachers in both
10 assessment scoring and reporting and all of these
11 things, but at the same time they have their own
12 job to do. And how are they going to actually get
13 involved in this very important, at the same time
14 needing a lot of training and a lot of background
15 to deal with these issues?

16 Not relying solely on the percent of
17 agreement, that's something that is not mentioned
18 that is easy. Inter- rate reliability. If we
19 just focus on percent of agreement, yes, it is
20 easy. But percent of agreement, as you know, is
21 the roughest and the least stable measure of
22 percent of inter-rate reliability. There are

1 several different other options, such as you know,
2 Kappa intra-class correlation and Williams' Index,
3 and many other aspects that are more robust, much
4 better approach to use, so if the consortia are
5 going to elaborate on the inter-rate reliability
6 ability approach, making sure that they address
7 some of the issues concerning some of these
8 computational and statistical issues, problems
9 with these things, and report or plan to report
10 more than one index of inter-rate reliability.
11 Because sometimes there are -- we have seen there
12 are major differences between different approaches
13 to inter-rate reliability in case of discrepancies
14 between these components, which one to trust more
15 and which one to do. Examine sources of
16 measurement error to inter- rate reliability.

17 So I think -- should I stop at this
18 point? My time is over?

19 One thing I wanted to mention, I made
20 some major changes, restructuring on my slides.
21 Not the content. I didn't change any of the
22 content so I have newer slides. The ones that you

1 have in your packet is not the most recent.

2 MS. WEISS: Yeah, no, we're posting your
3 correct version on the web.

4 MR. ABEDI: Okay. So should I stop at
5 this point?

6 MS. WEISS: Sure. It probably is a good
7 stopping point since we're just about out of time.
8 Let me see if there are -- are there any other
9 last -- I mean, is there anything you wanted to
10 just fast-forward through quickly and make sure
11 that you tell us?

12 MR. ABEDI: Yeah. Just one thing I
13 wanted to talk about the research -- I'm spending
14 a lot of time on running this.

15 So I wanted to go through the last -- I
16 think -- so let's see. I wanted to go to the
17 slide number 5.

18 SPEAKER: Two more slides.

19 MR. ABEDI: Let's see. Five. Okay.
20 Research evidence. Ideas that the Department has
21 for research are great ideas, but I believe some
22 of my colleagues mentioned that those ideas should

1 be part of the applications and applicants should
2 actually provide that. So I have a couple of
3 other ideas for research. One of them is a study
4 on how to make assessments more accessible for
5 ELLs and students with disabilities. If you
6 provide some guidance, it may help consortia of
7 states to be less dependent on accommodations that
8 may not produce valid results. Making assessments
9 more accessible for English language learners and
10 students with disabilities actually help to make
11 it more accessible for everyone because one of the
12 accessibility features that has been introduced --
13 we have introduced -- is making language
14 accessible and format accessible for these kids.
15 If you do that, then that's going to help assess
16 -- to help everyone.

17 Sources of measurement error, construct
18 irrelevant, differentiating affecting performance
19 of subgroups and ways to control the impact of
20 these assessments and also research on the
21 validity of accommodations used for ELL and
22 students with disabilities. That's one thing I

1 wanted to say. Some states or there are some who
2 need -- some people feel the need to provide
3 hundreds and hundreds of accommodations. There is
4 actually really -- there is no need to provide
5 that many accommodations. If you have research
6 evidence on a handful of accommodations for
7 English language learners, only using them would
8 be sufficient. There is no need to use many, many
9 of them. And research identifying some of these,
10 especially for English language learners because
11 their common needs would be assistance with their
12 language. So if you come up -- research points us
13 to some accommodations that are valid. It doesn't
14 change the construct being measured. Then
15 everyone can use it rather than wandering around
16 and using different accommodations that we don't
17 have any as far as any research evidence to
18 support it.

19 And then I don't know -- these are just
20 my own suggestions. It is important to
21 incorporate an ongoing quality control procedure
22 -- procedure by the Department into the process so

1 appropriate corrections can be made. I don't know
2 if that's possible or not. This is my dream.
3 But I don't know. Any departure or even minor
4 from the approved plans should be discussed with
5 the Department and justification for such changes
6 should be provided. The consortia should provide
7 assurance that the decision on the development and
8 implementation of the assessments are made
9 collectively by all the stakeholders. And then
10 lastly, it is important that people in charge of
11 developing the assessments are quite familiar with
12 the research findings in the area of assessment,
13 particularly with the recent (inaudible)
14 assessment for ELLs and students with
15 disabilities.

16 MS. WEISS: Great. Thank you.
17 Questions for Jamal? Oh, the post-lunch stupor.
18 If not, then let's pass -- do you have something,
19 Scott? Go ahead.

20 MR. MARION: Jamal, I was intrigued by
21 your comment on not getting carried away with
22 accommodations with ELL students and to just focus

1 on a handful. Do you think we know enough now to
2 know what fits in that handful or do we still need
3 more research to see what else is necessary? ELL,
4 not SWD.

5 MR. ABEDI: We have some basis. We know
6 -- we can identify a handful of accommodations for
7 English language learners and for some categories
8 of students with disabilities. But we may not
9 have enough research evidence to actually say for
10 sure. So we need some research to actually
11 identify, making sure that those accommodations do
12 not alter the construct. This is a validity
13 issue. It is extremely important that by
14 providing accommodation we don't give them unfair
15 advantage over others.

16 MS. WEISS: John, did you have a
17 question? Oh, sorry. Jeff?

18 MR. NELLHAUS: This is more of a comment
19 or just a suggestion of what you might have
20 applicants address in this particular RFP. And
21 that is how do we integrate the assessments for
22 ELL students -- the English proficiency

1 assessments with the other assessments in the
2 system? Because right now we have a lot of
3 testing going on with the ELL students. And the
4 question is can we get smarter about it, more
5 efficient about it? And that just might be
6 something you would want to consider having
7 applicants address in this process.

8 MS. WEISS: Thanks.

9 MR. EASTON: I'd like to ask a little
10 question about the quality control that you're
11 proposing. Do you see this different from some of
12 the things we heard before lunchtime about
13 internal research and developing evaluation
14 capacities within the consortia? Or do you see
15 this as separate?

16 MR. ABEDI: I see this as something
17 based on what you mentioned, as well as a group of
18 people who are experts in this area, including
19 Department personnel, to see exactly whether the
20 fidelity of the work on the application. Because
21 it is -- we all know -- we have all done research
22 and we know when we go to fill there are some

1 areas that we have to change our course of action
2 because we get to some issues at some point. So
3 in a case like that, what else -- what are the
4 issues? What else can we do?

5 MR. WISE: Good afternoon. Thank you
6 for the opportunity to come and talk further about
7 this exciting adventure that you all are
8 undertaking.

9 I apologize. I didn't realize I was
10 going to be last so you'll find a somewhat
11 redundant repetition of earlier recommendations
12 that were probably stated more clearly before.
13 And I also apologize that I'm also having problems
14 making this -- with the advance. Sometimes it
15 works.

16 I won't spend any time -- I had an
17 overview. It's the same five questions everyone
18 else was asked. So in the interest of time let me
19 just launch into them.

20 The through-course accountability
21 options I think are very intriguing. Like Scott
22 and others, I would urge you not to require the

1 method so much as to try and state clearly what
2 you hope the through-course assessment would
3 achieve. And then require the states to explain
4 how what they would propose would support that.
5 Now, a reason here especially is because I don't
6 think we have a lot of good models of how this is
7 done well. Now, we have a lot of good models. I
8 mean, most of us went through high school and
9 college courses where at the beginning the
10 instructor said half your grade is the final and
11 half is -- and a quarter is the midterm and a
12 quarter is the paper or whatever you're going to
13 write. And so there were components that went to
14 make up an overall summative score in evaluating
15 your work. And as sort of rather timid students
16 we never stopped to ask, well, where did you get
17 those weights? And question, you know, how this
18 all -- did it really fit together or so on.

19 There are at least three models I can
20 think of though about how the components might be
21 defined and how it might work. One is the
22 parallel forms notion. And I will point out that

1 actually Oregon -- and maybe some other states --
2 actually have multiple opportunities for students
3 to take virtually a parallel form of the same
4 test. So you're measuring the same thing, but at
5 different points in time. It gives you an early
6 indication. Maybe it gives you a better measure
7 of growth because you're not changing the measure.
8 It doesn't really provide an opportunity to test
9 in a lot more depth, which is one of the goals I
10 think some people have for through-course systems
11 because you have to cover the whole wide range of
12 content with each of the assessments. It also
13 requires you to test students -- if you're testing
14 early in the year, obviously they haven't had the
15 whole curriculum, but you're asking them questions
16 about everything, including things that they
17 haven't had as well as things that they had.

18 A second model is sort of the segmented
19 assessments. And this is -- in some sense this is
20 the end- of-chapter test model or its equivalent.
21 So you take the content, you divide it up into
22 pieces and you say, okay, I'm going to have one

1 good test of this piece and another good test of
2 this piece. And I hadn't thought quite this far,
3 but earlier this morning as we were talking about
4 maybe there's some things you can't assess in
5 real-time end-of-course and you have not just
6 different multiple tests, but portfolios or
7 components of listening and speaking or some other
8 things. But it should divide the whole content up
9 into segments. And you might then allow states or
10 districts or schools to administer the segments
11 whenever they thought they'd finish that piece of
12 the curriculum. So it wouldn't necessarily have
13 to be everybody is doing everything in the same
14 order. But they have to cover the whole nine
15 yards and they have to have tests of each of the
16 pieces as they finish those pieces.

17 One thing that that wouldn't do is it's
18 still sort of a one shot for the student. He
19 takes the test once. If he didn't get it then he
20 doesn't get a chance later on because you're
21 testing him on something else to demonstrate sort
22 of recovery from the early misunderstandings.

1 So a third model is the cumulative
2 assessment. You test the first segment of
3 curriculum in the first test. Then you test the
4 first and the second segments in the second test.
5 So you have opportunities to go back and retest on
6 things that the student might not have done or
7 comprehended quite fully on the first test and
8 show throughout progress, as well as covering all
9 the different segments.

10 If you're not careful, however, that'll
11 privilege what you taught first because it'll get
12 tested a lot of times and what you teach last gets
13 tested just once. So there may be some pros and
14 cons to these different models, but you need to be
15 clear about how you're designing the thing in the
16 first place. Will there also be a more
17 comprehensive end-of-course assessment that counts
18 the 50 percent or whatever rationale for weight.

19 I did want to stop and spend just a
20 minute because I thought we hadn't had the very
21 clear discussion and it's not clear in my mind
22 about the extent to which the consortia have to

1 agree not only on the common standards, but on a
2 common sequencing of the teaching of these
3 standards. And for some of these models they
4 would; for some of these models maybe they
5 wouldn't.

6 So the real question that we were asked
7 was not just to describe what they are, but to
8 talk about the kind of evidence that should be
9 required to demonstrate the technical rigor of the
10 assessments that are developed. And I must admit
11 my bias. I framed everything as validity. And
12 Scott mentioned this, everything is validity. But
13 there are many different types of validity
14 evidence that people collect that support
15 different claims or different potential uses that
16 are being made of these tests. So right now most
17 of the assessments -- we do content-related
18 validity. We do alignment studies. The question
19 is, is what's on the standards covered in the
20 test? And that's one kind of evidence. And that
21 would continue to be an important type of evidence
22 if we break the test up into parts and have sort

1 of through-course testing.

2 Correlation with other indicators of
3 achievement, you'd like to be able to show that a
4 teacher who spends -- the judgment of a teacher
5 who spends a lot of time with a student is not
6 inconsistent with the test results based on a
7 single or a couple of times. You might want to do
8 cognitive labs where you probe deeply about what
9 the student really knows and can do and then you
10 look at what the test results say and is there
11 some correlation or consistency there?

12 A newer area that we're not doing much
13 of now in the state assessments, but we really can
14 and should is more predictive evidence. I'm
15 pretty sure that a lot of states have gotten money
16 recently and probably turned in proposals
17 yesterday to continue to improve their
18 longitudinal data systems for student achievement.
19 And it's now much more feasible to ask at
20 different levels of success in one course, how
21 well the student did the next year or in the next
22 course that they took. And the idea behind the

1 Common Core Assessments is that we started with
2 the preparedness or readiness standards and we
3 hoped there would be a logical sequencing of
4 grade-by-grade standards that lead up to that
5 preparedness. So looking for evidence that the
6 assessments that you have actually build and
7 predict future success I think is a key and very
8 important new kind of validity evidence.

9 And then, of course, there's
10 consequential types of evidence that again it
11 varies with what the goals are for these uses.
12 But if the goals aren't -- and I will say that the
13 Department set out here's a list of goals we have
14 for this program and then here's a set of
15 questions. But they didn't draw the one-for-one
16 map and say, well, now, this question is really
17 about how we get to these goals and others. So I
18 think you can ask the states as they're responding
19 to their Request for Proposal to be clear about
20 how different pieces of validity evidence will
21 support different goals. So impact on learning
22 instruction and improving learning instruction

1 more than just holding schools' accountability is
2 a high level goal. And you'll need sort of
3 surveys and observations of the curriculum of
4 instruction to see what actually happens as a
5 result of this sort of through-course assessment
6 to tease out how the through-course elements
7 really either enhanced or failed to enhance the
8 impact -- the positive impacts on instruction and
9 learning. And as we heard a little bit this
10 morning, either prevent or reduce the negative,
11 like the narrowing of the curriculum aspects.

12 And the finally, obviously, you want to
13 look at student achievement over time and answer
14 the ultimate question -- are students learning
15 more now with these new, better assessments and
16 the integrated systems than they did then?

17 Okay. So I -- on the common high school
18 end-of- course exams I was struggling with sort of
19 what's common. Common across what? And the first
20 area of commonality was sort of across different
21 subjects are we having tests that are equally
22 rigorous and so on? And I think this is very

1 challenging. Because take, for example, you had
2 one set of assessments designed to evaluate
3 student performance in physical education class
4 and another in an Algebra II. And at what level
5 are they comparable with regard to difficulty or
6 rigor or so on? So it's a very potentially broad
7 challenge.

8 And a way to meet this challenge is to
9 not sort of focus on just normative data -- well,
10 they're equally rigorous because 50 percent of the
11 kids pass in each course. But to look at, first
12 of all -- first of all, ask that there's a common
13 process. And it's usually experts are brought in
14 and they judge. And to get to the agreement on
15 the content standards because really the tests
16 have to reflect what's in the standards that the
17 tests are trying -- so the commonality is --
18 should be more focused on the content standards
19 than necessarily on the tests. Although in
20 subsequent discussions, if you were to want to
21 allow school- based tests and some wide arrange of
22 tests, then the comparability issue does become do

1 these different tests actually have comparable
2 rigor in covering what's on the common content
3 standards?

4 Now, one of the values in having a
5 common test, even if you have somewhat difficult
6 curricula -- so the thing is, well, would you have
7 to teach exactly the same? Well, you don't have
8 to. If you've got agreement on what the goal
9 should be, the content standards, and you've got a
10 measure of how well those are being accomplished,
11 this is a great way to try out and evaluate
12 different curricula, different orders, different
13 ways of teaching the same things because you've
14 got this sort of common measure across the states
15 and the consortia of what it means to succeed in,
16 you know, a high school biology course or so on.

17 I also, you know, want to talk about
18 another use of these high school course
19 assessments. And that is high school
20 accountability. Right now -- and we don't know
21 what reauthorization will say. This may be as
22 much about reauthorization, but right now the "and

1 once" in the high school years is the requirement
2 that's implemented very differently in many states
3 and frankly, I think most of the time is a very
4 inadequate indicator of what high schools really
5 contribute to student learning. And a reason why
6 this is difficult is because by the time they get
7 to high school students branch out and they learn
8 different things by design. And so we need some
9 way of having comparability across the different
10 things that they learn to say sort of what value
11 is being added for each student and is there some
12 way to sum those up?

13 So some of the state graduation tests
14 are an example of how this school level
15 accountability could be structured. You have a
16 core set of courses. Maybe it's just English II
17 and algebra or it's English II and algebra and
18 some general science. And then you have a set of
19 sort of electives and you have to pass all of the
20 core and maybe up to three or four -- some number
21 of electives. And then you could ask for the high
22 school, what percentage of their students is

1 succeeding and meeting those requirements? And
2 that would be a way of getting high school
3 accountability that's broader, that recognizes the
4 need to teach a lot of different things -- not
5 just one or two things -- would help maybe avoid
6 the narrowing the curriculum particularly at the
7 high school level. And it would be then really a
8 better added of the value-added -- better model or
9 better indicator of value-added at high school.

10 Okay. A third area is challenges for
11 computer- based testing. And, you know, a lot of
12 us used to complain about Microsoft products
13 because they came out and they were real hogs.
14 They required all this memory. It seemed really
15 like exorbitant requirements. And then a year or
16 two later it seemed like they were trivial
17 requirements and we don't talk about that anymore.
18 And I think the design of this program needs to
19 take sort of the same approach. Don't worry too
20 much if, you know, design for what's the best that
21 we can currently imagine. And unless there's a
22 real stoppage in the technological development,

1 that will be reasonably common place and trivial
2 by the end of the four-year grant.

3 There's -- one of the questions is about
4 the comparability across modes or comparability
5 that we're supposed to address. And the concern
6 is that we not be limited by -- just because we
7 can't do this well in paper and pencil we
8 shouldn't do it on the computer. So,
9 comparability I think needs to be asked, you know,
10 comparability for what? And it comes up in the
11 context often of accommodation. So for students
12 that can't take the test on computer, is there a
13 paper and pencil option? Well, in the case of
14 accommodations, accommodations aren't supposed to
15 give comparable scores for every student. They're
16 supposed to give appropriate scores for students
17 who need them. And so I wouldn't get hung up on
18 huge comparability studies of paper and pencil
19 versus computer in order to answer the question
20 are these appropriate tests or modes for the
21 students that need them.

22 And one thing I would point out is that

1 universal design principles apply to computer
2 tests as well. So if you're finding things aren't
3 comparable, you have to ask is the reason that
4 they're not comparable because you're measuring
5 things that are irrelevant to the construct? You
6 know, are you requiring things of the student that
7 have nothing to do with what you're trying to
8 measure? And then get them out of there to start
9 with.

10 You know, I took a few notes on things
11 that I wanted to be sure and add and I can't --
12 oh, yes. One other thing about the computer-based
13 testing. Test security becomes an issue that
14 people will have to address and say how they're
15 going to address. It's one thing if you're giving
16 a test to all the students at once, which you can
17 do fairly easily in paper and pencil. If you have
18 to test a few students at a time over an extended
19 period of time and you have stakes attached,
20 either for the students or the schools, you know,
21 the thing that almost sunk the GRE is not
22 realizing sort of how much broader the item pool

1 needed to be in order to prevent issues related to
2 test security.

3 The fourth one I've labeled continuous
4 process improvement. And I'm a big fan of
5 continuous process improvement. I think the odds
6 that a state or anybody will get this exactly
7 right on the first try are about zero. So you
8 need to have a process in place, both during the
9 grant and certainly for sustainability beyond the
10 grant, to continue to improve the process by which
11 the test is developed. Maybe even improve the
12 common content standards that people are agreeing
13 to.

14 So, some examples of how this is done is
15 looking at individual test questions. And when
16 they don't work well in trial, you know, not just
17 saying, okay, they didn't work well. We'll throw
18 that aside. Let's go on. But also saying, okay,
19 why didn't they work well and what could we have
20 done to revise our item writing procedures, our
21 item writing guides, our item review guides, in
22 order to have weeded these out much sooner and to

1 have a, you know, higher set of quality items.

2 If we're trying to evaluate the impact
3 on teaching and learning, we need to look at sort
4 of what schools are doing in response to this and
5 then ask is that what we wanted? And what could
6 we do by changing the content, the mode of
7 instruction, the use, or whatever in the process
8 to further enhance its positive impacts on
9 teaching and learning and reduce or avoid negative
10 ones?

11 And then the predictive powers of the
12 assessment is something where there's a potential
13 for a continuous stream of feedback information
14 that can be used to go back and say what could we
15 do to improve the degree to which the standards
16 and the assessments at one grade are aligned to,
17 predictive of, useful for screening students'
18 development with respect to what happens later.
19 So there's a need for an ongoing developed test
20 and revised cycle sort of throughout the whole
21 system. And states would be sort of, I think,
22 well advised to be clear about how those cycles

1 will work.

2 One aspect that's commonly used is
3 Technical Advisory Committees. And I would think
4 states would be well advised to think about how to
5 most effectively use Technical Advisory
6 Committees. I mean, those of us that serve on
7 them, you know, when a state actually pays
8 attention to what they say and actually does
9 something, it may be more valuable than just
10 hearing us talk.

11 Further research --

12 MS. WEISS: Not that we're not enjoying
13 that.

14 (Laughter)

15 MR. WISE: Well, actually, in one of the
16 states that I work with that shall remain
17 nameless, somebody new came in and at the
18 beginning of the next meeting they went through a
19 whole half hour of what they had done in response
20 to what we said last time. And then we got very
21 serious about being careful about what we said
22 after that. Well, so there are several areas -- a

1 couple of areas where more research is needed.
2 And I would certainly agree value-added and
3 performance tasks, like Scott, I would focus more
4 on what you want to achieve and make sure that
5 you're not just saying here's a great method.
6 What can we do with it? So if the goal, for
7 example, is to really improve how we coach,
8 evaluate, and improve teacher effectiveness, then
9 sort of look at it from that lens specifically and
10 not just -- and ask, you know, are there models or
11 even different ways of doing that?

12 Certainly with performance tasks there's
13 been a lot of work on that. There's some
14 problems, especially on equating. And there's
15 challenges, I think, to try to assess sort of more
16 complex skills, inquiry problems, or team skills
17 that aren't really accessible easily now in
18 multiple choice. I would add the preparedness and
19 prerequisite skills are an areas where we're still
20 really experimenting. We need to look at how
21 those are specified. And then to look at the
22 learning trajectories that lead to those skills.

1 And I think Scott mentioned this as well. But
2 this is an area for continuous improvement and for
3 sort of deeper sort of cognitive research on how
4 this all fits together.

5 So, in the 20 seconds I have left,
6 through-course assessments -- I would really
7 support using the cumulative or the segmented
8 models to increase coverage of the different areas
9 in greater depth that require multiple types of
10 validity and impact evidence. The high school
11 end-of-course exams I would use -- I would really
12 support using them as part of more improved high
13 school accountability measures. Challenges for
14 computer-based testing, well, you can read the
15 words here because I see the red light. And I
16 quit.

17 (Laughter)

18 MS. WEISS: Questions for Laurie?

19 Thanks, Laurie. That was terrific.

20 So I have a question going back to your
21 slide on the continuous process improvement. And
22 you talk about analyzing the predictive power of

1 assessment results and say that one of the things
2 we should -- applicants might look at are some
3 essential skills not covered? Could you unpack
4 that a little bit? What are the kind of questions
5 that people would be asking and researching to
6 answer that question?

7 MR. WISE: Well, you know, an advisor is
8 intentionally vague so you can read more into it
9 than we actually realized up front.

10 There are a couple of examples. So one
11 is in the through-course assessments you have
12 different pieces. You have different weights.
13 Some of them might do a better job at predicting
14 how well the student did the next year and
15 therefore might be more critical than others. And
16 you might want to tweak and adjust the weights.
17 You may have some students who do very well at one
18 grade and then they don't at the next. And so you
19 want to look at what skills did they lack that
20 caused them not to succeed at the next grade. And
21 could you have anticipated or incorporated those
22 into the measures at the earlier grade so that you

1 could identify and remediate at a much sooner
2 point?

3 MS. WEISS: Other questions for Laurie?
4 Go ahead, Jeff.

5 MR. NELLHAUS: No, I'd just like to -- I
6 like the way you set up the description of the
7 different kinds of through-course assessments.
8 And you have the parallel forms, the segmented,
9 and then the cumulative. And I think another
10 additional one, one I was referring to earlier,
11 might be the targeted model where you're just
12 targeting particular standards once or twice
13 during the year rather than thinking about a whole
14 assessment of all of the standards. That we're
15 really targeting certain standards. So it might
16 be just another option as you think about the
17 through-course assessment system.

18 MS. WEISS: And you're thinking about --
19 give an example of what you mean by that? What's
20 something you might choose to target or what's
21 something you might choose not to target?

22 MR. NELLHAUS: Oh, it could be an area

1 of English Language Arts. It might be a listening
2 test. Or an area of Science, it might be a
3 scientific investigation of some sort. So
4 something that's, not again, doesn't lend itself
5 to the on-demand setting, but could be given
6 during the year and thought of as a type of
7 through-course assessment.

8 MS. WEISS: Thanks. Any other
9 questions?

10 MR. ABEDI: One issue that -- I mean, I
11 agree with what Laurie said, but one of the major
12 issues that I absolutely (inaudible) was part of
13 my presentation, but I didn't get a chance to -- I
14 should have managed it better -- was the
15 comparability between computer -- on paper and
16 pencil test that it is almost for some cases
17 impossible to actually establish that
18 comparability because there are some capabilities
19 that computer-assessment has that paper and pencil
20 may not have. So in that sense it is not possible
21 to actually think or talk about comparability
22 because the areas that the computer system

1 actually adds and improves the construct may not
2 be possible in paper and pencil. So that's --

3 MS. WEISS: Right.

4 MR. WISE: I did want to add one more
5 thing on continuous process improvement that
6 builds actually off Jamal's
7 suggestion/hope/wish/question and that is the
8 Department needs to think about what kind of
9 flexibility it will allow for the work plan to be
10 altered as the grant goes along. And, because,
11 again, rarely do we have the design just right at
12 the very beginning. So whether you have sort of
13 annual updates that you will negotiate or whether
14 -- you know, and how much is this really a grant
15 versus a contract where it's tightly specified and
16 they have to live to, you know, the details of
17 what's being proposed.

18 MS. WEISS: Okay, so should we -- so
19 since we ended a minute ago with the summative
20 course assessment notion, let's start there. And
21 I think this question came up earlier, but let us
22 also give credit for the term to one of our

1 presenters that we had a couple of months ago. I
2 think Linda Darling-Hammond is the one who coined
3 this term of through-course assessments. And I
4 think there was a pretty specific idea she had in
5 mind for it that was around performance tasks or
6 maybe even what Randy Bennett and Drew Gitomer
7 called in their paper "foundational tasks" that
8 were given periodically as appropriate throughout
9 the school year as sort of big tasks that were
10 then somehow based on whatever weights people
11 chose to ascribe to them aggregated together with
12 other types of assessments and item types that
13 were given either throughout the year or the end
14 of the year to come up with some kind of summative
15 score.

16 And the question, I mean, we put it out
17 there as should it be required partly to just
18 force a real conversation about this so that we
19 could understand because I think we think that the
20 notion has some real benefits. We're not sure at
21 all whether it should be required or just
22 something that we should talk about. And one of

1 the things that you might be able to help us
2 understand better is what are the tradeoffs
3 between -- what are the pros and cons of doing
4 some kind of through-course assessment versus a
5 comprehensive end-of-year assessment? And are
6 there ways that as states are thinking about this
7 and responding to the RFP, assuming we don't say
8 it's required, but it's optional, are there
9 reasons or grade levels or content areas or it's
10 always applicable -- are there places where it
11 might be more or less applicable? Where it makes
12 more or less sense? How do you guys think about
13 after these conversations the pros and cons and
14 tradeoffs across those two types of assessments?

15 Yeah, just jump in.

16 MR. BENNETT: I would frame my answer to
17 that in terms of three premises. The first
18 premise is that accountability assessment isn't
19 going to go away. It's not going to go away
20 because policymakers need it, and they need it at
21 all levels.

22 The second premise is that how we do

1 accountability assessment has a profound impact on
2 what it is that happens in classrooms. It affects
3 the way teachers teach; it affects the way
4 students learn.

5 And the third premise is that if we can
6 get it right in the sense of create assessments
7 that have better -- more positive impact on what
8 it is that students and teachers do and that
9 provide accountability information, we'll have
10 provided an important service and made an
11 important advance.

12 One reason to advocate through-course
13 assessment is the potential that it could have
14 that impact if designed appropriately. It's not
15 going to have that impact if it's simply comprised
16 of a series of multiple choice assessments that
17 are given throughout the school year. It might
18 have that impact if we can create it from tasks
19 that are more like the ones that we want students
20 to be able to do in real world settings that more
21 closely replicate what it is that proficient
22 performers have to do to succeed in a domain. And

1 if we can create assessments that have that
2 character and distribute them throughout the
3 school year, I think we'll be moving in that
4 direction.

5 The other side of this argument is that
6 you can't do that in a single end-of-year
7 assessment. That is, you can't create a single
8 end-of-year assessment out of such tasks because
9 you won't be able to put enough of them into that
10 single end-of-year assessment to create a valid
11 generalizable measure.

12 MS. WEISS: Yeah, Scott.

13 MR. MARION: Just to pick up an actually
14 question Randy a little bit on this and then to
15 perhaps move it a little further along -- so take
16 Randy's first two premises. Accountability isn't
17 going away and how we do accountability has a
18 profound impact I think has considerable
19 implications for the notion of being able to get
20 it right. We could design the best assessment
21 ever designed. And if we put it in an
22 accountability system we could easily have that

1 distorted so badly that we wouldn't get nearly the
2 benefits we hope to get out of it.

3 So, what I would like to see -- well,
4 we've known each other long enough so you know
5 obviously I support the kinds of assessments
6 you're talking about. But I think that one of the
7 constraints that we have to be aware of and the
8 consortia proposals need to address this to the
9 extent that they can given the information or lack
10 thereof about the reauthorization, is thinking
11 about how these will get used within the
12 accountability system that will allow them to
13 realize the benefits that I think you're hoping
14 that could be realized.

15 So, for instance, if the teacher is just
16 the test administrator for these through-course
17 assessments, then we might as well not bother. I
18 mean, or do something different because then
19 they're just -- it's just another deskilling of
20 the teaching profession. And they do this and
21 they pass the results on. And then somebody will
22 tell them how their kids did. And I think we need

1 to really think about changing the model a little
2 bit. And by loosening some of the accountability
3 constraints around how these things -- how the
4 whole assessment system gets used. But these
5 pieces particularly I think we could -- so for my
6 -- if I was the one saying what are my goals for
7 the system, it's to allow kids to wrestle with
8 meaningful content in profound ways. And so they
9 could further their potential and skills for
10 whatever they want to do after high school. Then
11 I would want them to be able to have the
12 opportunity (inaudible) not always accountability
13 settings. And I don't think you were pushing for
14 the (inaudible) accountability settings.

15 MS. WEISS: Well, but let's push that
16 because maybe Randy kind of was. Maybe Randy was
17 saying if we don't put it in the accountability
18 system it'll get done by the schools that are
19 already high performing and doing it and it won't
20 get done by the others. And in the end we won't
21 have solved the equity issues that we're trying to
22 address here.

1 So how important is putting it into the
2 accountability system versus will the
3 accountability system pervert it so we can't --
4 like, those were sort of two diametrically opposed
5 premises there.

6 MR. MARION: Let me just -- I'm not
7 opposed to having accountability, but I think some
8 of the narrow ways we've been thinking about
9 accountability has limited the potential. So
10 there's lots of accountability in other countries
11 around these through-course and end-of-course and
12 performance tasks that has meanings for kids, has
13 meanings for teachers, but doesn't necessarily
14 have to be considered in the same exact way we're
15 doing it now where everything has to be the same
16 for everybody at the same time.

17 MS. WEISS: Yes, Jeff.

18 MR. NELLHAUS: Just building on what
19 Scott just mentioned. I think what we lost here
20 is the word -- we're trying -- we're using this
21 phrase now "through-course assessment." And the
22 way I used to think about it until you introduced

1 these questions was curriculum-embedded
2 assessment. And the difference there is that with
3 through- course it's sort of this external test,
4 supposedly -- at least the way it's been talked
5 about -- coming in. And it's just another
6 summative-like test that's given throughout the
7 school year. Curriculum-embedded is something
8 that's tightly aligned with the instructional
9 program. It's something that students don't --
10 you know, it's kind of part of the daily classroom
11 activities that somehow is collected.

12 So if we're thinking about writing and
13 we're thinking about students that are normally
14 writing compositions and different kinds of pieces
15 during the school year, that somehow that's
16 standardized a little bit. There's some rubrics
17 for scoring it and that information could get
18 collected and scored by the teacher, with training
19 and support and some auditing, and get factored
20 into the English- language arts score. That's
21 very different than thinking about, okay, it's
22 December 1st. It's time for the through- course

1 assessment this year.

2 MS. WEISS: Right.

3 MR. NELLHAUS: So, you know, I think we
4 can think about through-course as -- if we can
5 think about it as curriculum-embedded as well, I
6 think I'm more comfortable with it.

7 MS. DeSTEFANO: And I think another
8 dimension of that is it's complementarity with the
9 end-of- course assessment or the end-of-year
10 assessment. So to use the through-course to get
11 at things that are difficult to assess in that
12 standardized summative assessment and to look at
13 formats and modes of assessment that are different
14 than the standard, multiple choice objective
15 assessments. So states are considering it as a
16 system and then thinking how could these embedded
17 pieces reflect good instruction and also expand
18 what we learned from the end-of-course
19 assessments. And I would argue if it's not
20 expanding what you're learning, then it's not a
21 good use of through-course.

22 MS. WEISS: Yeah, go ahead.

1 MR. WISE: I also think that we have a
2 great place for, you know, high fidelity,
3 interesting performance tasks that will drive
4 instruction in a positive direction. But I think
5 that in areas where the content requirements or
6 the skills -- the standards we're hoping to
7 achieve -- require it. And it's not just because
8 this is cool.

9 So it may not be that in every area you
10 would necessarily need to have -- to assess things
11 that can't be assessed with the current
12 end-of-course assessment. So, in science, I'm
13 convinced inquiry skills is an area where people
14 have made a pretty good case -- benefit from a
15 much richer, more elaborate form of assessment
16 than just -- than we have now. But I'd caution us
17 not to try to over-generalize that every subject
18 needs these kinds of tasks as part of the
19 assessment.

20 MS. DeSTEFANO: I would agree with that.
21 And I think that Jeff's example of language arts
22 where we're constantly challenged -- how do you

1 assess listening? How do you assess speaking?
2 That's a nice place. The idea of a scientific
3 inquiry that can occur over time, maybe be
4 collaborative. So, again, well chosen that
5 complement things that we can't generally get from
6 the end-of-course.

7 MR. ABEDI: I think the idea of
8 through-course assessment is very promising. And
9 I think it's very important. The concept -- the
10 thoroughness, comprehensiveness involving
11 teachers, involving other modes of assessment
12 (inaudible) -- I mean, performance assessment.
13 And getting teachers involved. I'm so glad that
14 Linda Darling-Hammond is in the audience today
15 with us. She can direct us on how to get teachers
16 involved, performance assessments, and all those
17 things and stuff like that.

18 But just a word of caution that not to
19 rush on making the requirement because we don't
20 have enough evidence right now -- enough
21 information. What I would suggest is maybe a
22 series of focus groups with teachers, with school

1 principals, with educators, with others involved
2 to see. I'm not saying that not at this point to
3 do this. Do this, but in order to make it a
4 requirement, wait until we get some better
5 information.

6 MS. WEISS: Do you guys have any more
7 follow-up questions on this one? Ann or -- go
8 ahead.

9 MS. DeSTEFANO: I think another area --
10 we've talked about getting teachers engaged. I
11 think this is another area where we should look at
12 student engagement. I don't know that many
13 students that get hyped up about taking
14 end-of-year exams. I can think of some schools
15 where they had these traditions, a senior essay, a
16 junior science project, where it becomes part of
17 the school tradition and the students are very
18 engaged. So I think student engagement should be
19 a piece of examining the use of this type of
20 assessment as well.

21 MS. WEISS: So can I just go back and
22 re-ask the question of is there a reason this

1 needs to be folded into the accountability system
2 even if its weights are different versus is there
3 a reason it should not be? In other words, I'm
4 worried about the question that came up earlier
5 about are we crowding out interim and maybe -- I
6 mean, formative I think of as different. I think
7 of that as really sort of daily embedded totally
8 in the lesson structure, but are we crowding out
9 interim? Is that a bad thing? Or are we putting
10 better things in place of current interims? And
11 for high performing schools that may crowd them
12 out, but in most schools it's good. How do you
13 guys think about that?

14 MR. WISE: Hopefully replacing interim
15 because the idea that you do interim testing and
16 three-quarters testing and just more and more
17 testing. And interim testing varies a lot in
18 quality now.

19 MS. WEISS: Right.

20 MR. WISE: And in design. And it's not
21 always clear. Is it just going to predict the
22 end-of-course score or what? So a single, well

1 thought through system has got to be better than
2 two half thought-through systems.

3 MR. NELLHAUS: Yeah, so Joanne --

4 MS. WEISS: It sounds like a low bar,
5 but actually it's not. That's the sad thing. Go
6 ahead.

7 MR. NELLHAUS: So if you think about
8 this through- course as being a series of three or
9 four tests, you might think of them as benchmark
10 assessments that would be optional. You know, so
11 they wouldn't count towards the accountability,
12 but they would be -- we would get high quality
13 benchmarks this way. And for schools that were in
14 serious accountability categories, the state could
15 require that they be administered. So it would be
16 more of a tool for those schools that need to look
17 more closely at data to make improvements to do
18 so. So, you know, the notion of a benchmark
19 assessment is very similar to this, only we
20 weren't thinking of rolling up the benchmark
21 assessments into the summative score. That's the
22 only real difference I see. So I think there's a

1 real role for this. You know, the question is do
2 they get used for accountability or not? And if
3 not, then they can be used for these other
4 purposes which might be equally as useful.

5 MR. BENNETT: I think you would want to
6 use them for accountability, but only if you could
7 create assessment that effectively modeled good
8 learning and instructional practice that had a
9 positive impact on what it was the teachers and
10 students were doing in the classroom. And that
11 gave you accountability information that was at
12 least as good as what you were getting now.

13 MR. MARION: Could I just add one other
14 thing to this? I actually don't think we know.
15 And so I would actually urge, again, to beat this
16 horse, to be really clear about what we're trying
17 to do. And then even within a consortium -- say
18 in Randy's state they determine that they figured
19 out a way that they want to use these -- and I'd
20 much rather see Jeff's term -- no offense, Linda
21 -- of curriculum- embedded because I think you
22 actually resonate with the idea a little more, too

1 -- because to pretend that it's not related to
2 curriculum is a lie. And so whether we call -- I
3 think these curriculum-embedded assessments as
4 part of the accountability system -- and I could
5 certainly think of ways that I would roll them out
6 for schools, for students, for whatever it might
7 be. And in other places where it is sort of hands
8 off on stakes. And allow some experimentation
9 within because I don't really think that we know.

10 MS. WEISS: Go ahead, Judy.

11 MS. WURTZEL: So going back to Randy's
12 point that one of the reasons to think about
13 through-course assessment is the opportunity to
14 have well designed tasks that ask students to
15 demonstrate what they know across multiple
16 standards to generalize more complex work that's
17 harder to measure on an on-demand context. So if
18 you decided not to do through-course assessments,
19 how much of that could you gain in an
20 end-of-course assessment and how much of that
21 would we be losing from our accountability system?
22 What are the implications for how you start to

1 think about your end-of- course test --
2 end-of-year test?

3 MR. MARION: As long as the end-of-year
4 started March 1st you could probably do all the
5 same, but you can't. You've got to give stuff up.
6 I mean, I think Laurie's science inquiry or
7 science experiment case is, you know, a perfect
8 example. You just can't do that. I mean, you
9 can, but you can't do enough of them to really get
10 the kind of generalizable evidence that people are
11 talking about.

12 But, again, I think it doesn't mean
13 because it's not being done and required by the
14 state or by the consortium that it's not being
15 done. And so the question really becomes if we
16 had this common sort of summative component -- and
17 in Randy's state they're doing it. They're doing
18 it as part of the system. And in my state it's
19 voluntary. And lo and behold we have about the
20 same trends. Or we find that in Randy's case
21 they're growing at a much faster rate and
22 demonstrating more knowledge and skill. So we

1 have a chance to actually learn something about
2 what might be the most appropriate way. And I'm
3 guessing -- I'm more than guessing here -- that
4 it's going to differ by context. And it'll be an
5 interaction of sort of school quality and
6 usefulness of this. The top schools don't need
7 you to tell them what to do to prepare their kids
8 to move on to the next phase.

9 MR. WISE: I think we have a nice
10 conundrum posed here that -- making something part
11 of the accountability system, so giving weight to
12 the through-course test -- is that a good or a bad
13 thing? It's good if it focuses people's attention
14 on good models of how to do things. It's not good
15 if people then treat it in a different way and
16 narrow the curriculum. It has a negative impact.
17 And we don't know in general or for a particular
18 subject. So maybe we can get John Easton to help
19 fund some research -- I got your attention -- that
20 would actually begin to address this question more
21 systematically than any one state or consortia of
22 states can do within the context of this grant.

1 MS. WEISS: Any other questions before
2 we turn to the high school?

3 MS. WHALEN: I wanted to kind of go back
4 to the comment of it being curriculum-embedded
5 because that has very different consequences for
6 what it means to be part of a consortium then if
7 that is the thinking behind this. So I just
8 wanted to get people to delve a little deeper on
9 it because it's easy to say that off-the-cuff. I
10 think it's hard to actually implement that across
11 -- within the state, let alone across multiple
12 states. So if we could just talk -- maybe if
13 someone has some ideas about the pros and cons of
14 having -- what has to be common for that to work?
15 And then whether that really is something that
16 should be or could be part of a real design as
17 part of a common assessment.

18 MR. MARION: Good question.

19 MR. NELLHAUS: I think the, you know, to
20 the extent possible, the tasks need to be common.
21 And training on how to administer them and how to
22 score them need to be common. Now, if you don't

1 get common tasks, then you have to have criteria
2 for the development of tasks that address certain
3 standards in a particular way. So, again, it gets
4 to just how much standardization you want. But if
5 you're looking for some comparability, you know,
6 you're going to have to standardize the tests.

7 MS. WEISS: I think Ann's asking whether
8 you have to standardize the curriculum.

9 MR. WISE: Curriculum.

10 MS. WEISS: I mean, that was why we
11 liked Linda's words better because --

12 MR. NELLHAUS: Well, you develop tasks
13 that are based on the standards. And those
14 standards should be embedded in the curriculum.
15 You know, and when schools are addressing those
16 standards might vary. But, you know, you're
17 developing tasks that, you know, most good
18 teachers and good schools would be developing on
19 their own to address the standards. So, you know,
20 you begin to work with the field to identify tasks
21 that seem reasonable for that curriculum and those
22 standards.

1 MR. BENNETT: Yeah. I think I agree. I
2 don't think you need to standardize the
3 curriculum. What you need to presume is that
4 there's going to be a set of core understandings
5 or big ideas, if you like, that are going to
6 appear in whatever curricula are being used in
7 that consortia. And that these tasks or periodic
8 assessments or through-course assessments or
9 whatever you want to call them are going to target
10 those core understandings or big ideas. And,
11 further, that those core understandings and big
12 ideas are going to occur in some sequence that you
13 can target the assessments to.

14 MR. WISE: I would also say I think
15 we're hoping, most of us, we're not talking about
16 something that's given every week as part of the
17 through-course assessment. So we're not having to
18 have commonality of what happens across states at
19 that level. And there are different models for
20 how the through-course might work. The segmented
21 model where the states can give it in any order
22 they want is one that doesn't push a common

1 curriculum. You do have to -- I mean, the big
2 sort of gain is you've got states going in and
3 agree to the common objectives for the course as a
4 whole.

5 MS. WEISS: Right.

6 MR. WISE: And then other models may be
7 some agreement on the prescribed order of the
8 major pieces of those objectives without
9 necessarily getting down into exactly what's on
10 page 34 of the curriculum.

11 MS. DeSTEFANO: And an agreement on
12 those skills and knowledge that are typically hard
13 to test through traditional means.

14 MR. MARION: Just a slight dissention.
15 I think if it's done the way Randy sort of said
16 quickly -- the focus to each of these tests are
17 focused on big and important ideas of the
18 discipline at that age or grade level, that's one
19 thing that might get some agreement. But I think
20 curriculum matters. And I don't think the
21 standards -- no matter -- I'm sure they're going
22 to be the best thing we've ever seen when we see

1 the final draft, but -- complete confidence in
2 that -- but I think it still won't be enough to
3 guide the day-to-day teaching. And there will be
4 considerable variability by quality of teacher, by
5 quality of school, in how well those get
6 implemented.

7 And we see it now. I've been doing this
8 for a while. There's a wide variety of
9 implementation. Supposedly everybody in the state
10 is teaching the standards now and we know that
11 looks very different. And a lot of times people
12 just buy textbooks and they just hope it's close
13 enough. And so I would love to see -- I'm not
14 saying you have to specify the exact curriculum in
15 a consortium, but if I was going to do this as a
16 consortium, I would want to at least know the
17 narrow -- the range of what's possible. Because
18 we know that if we pick curriculum and design
19 curriculum A, B, or C, it will at least have
20 enough of the big ideas. If you want to implement
21 A and I implement C, that's fine. We get a good
22 test. But I also think these performance tasks

1 will exacerbate differences in opportunity to
2 learn or reveal bigger differences if people don't
3 have at least some common ground.

4 So I don't totally disagree with
5 everybody. I just want to --

6 MS. WHALEN: So I want to push you on a
7 little bit -- something, Scott, is that previously
8 in other panels -- and I believe in the one on
9 Boston that you were a part of -- that there was
10 kind of this overwhelming suggestion that we
11 really think about systems of assessment and how
12 they fit together and how pieces fit together.
13 But I've heard a couple times here today you
14 talking about individual states or localities
15 dissenting from pieces or implementing them
16 differently as part of a system.

17 So I just wanted to kind of help -- you
18 help me wrap my brain around -- or if other panel
19 members have kind of this -- we need a system of
20 assessment, but then having different pieces be
21 voluntary versus required. And then how it
22 actually rolls back up into a real system and how

1 then you have a theory of action as a consortium
2 coming together if there's so much dissent and
3 variation.

4 MR. MARION: Yeah. I wouldn't call it
5 dissent and variation. I would call it trial and
6 experiment. Because I don't think we know. And
7 right off the bat --

8 MS. WEISS: Oh, well, then we're sold.

9 SPEAKER: It's all in the words.

10 MR. MARION: That's all that matters.
11 To be honest, I think you heard correctly in
12 Boston and I think thinking about the constraints
13 and the legal constraints that you guys are under
14 and the requirements for getting this money out
15 the door -- I know it's a fast timeline. And
16 thinking about work and states and talking with
17 certain sort of consortia forming and unforming --
18 rethought this and think that really if innovation
19 is the goal, then I think we've got to figure out
20 something we could do. And it might look
21 different within a consortium, but there would be
22 enough common pieces within a consortium that it

1 would pay to work together.

2 So I didn't come up with the idea of
3 consortium. I think there are certain things
4 about it that are advantages, but they also create
5 extra work. So I think there needs to be an
6 overarching umbrella system that's common, but I
7 think within the consortium I think we could learn
8 a lot if people had that common umbrella, but then
9 did different things. So a few states really
10 worked hard on innovations in computer- adaptive
11 testing in early math and others did something
12 with really great performance tests in listening
13 and things like that that I think there's ways --
14 but it still rolls up and eventually gets shared.
15 So I don't think you heard that differently in
16 Boston. I think we've evolved a little bit, as
17 you have.

18 MR. BENNETT: Yeah. I think the idea
19 behind system was to have a set of components that
20 worked together in a synergistic way. And there's
21 nothing about the idea of a system that precludes
22 having optional components, but those components

1 ought to be designed so when they're used they
2 work in synchrony with everything else.

3 MR. MARION: I love sitting next to
4 Randy. He's like the smart kid in the class.

5 MR. NELLHAUS: And the system doesn't
6 mean that we take all the different information
7 from all the pieces of it and roll it up into a,
8 you know, one single score either.

9 MS. WHALEN: No, no, no. And I didn't
10 assume that.

11 MR. NELLHAUS: Okay.

12 MS. WHALEN: One thing I did assume is
13 that one of the values of having through-course
14 assessments is that you then don't have to sample
15 all of the standards at the end -- a summative
16 course. And so then there has to be this approach
17 in theory that you can then sample throughout the
18 year and have that then get a comprehensive
19 picture of what a child has mastered and is able
20 to do or what he or she hasn't and be able to form
21 instruction and think of intervention and
22 strategies. And if that becomes a voluntarily

1 component, you're missing a big piece of assessing
2 what that child is able to do to inform what your
3 next steps are. So I don't know how you could
4 have a theory of action or a comprehensive system
5 that says this is how we're going to approach
6 sampling of standards and then be able to have
7 that be something that individual states opt to
8 innovate or research versus others not and then
9 still say that this is a system that's designed
10 and measuring and implementing what we say it's
11 going to do.

12 MR. MARION: Yeah. I mean, I think you
13 raise some great questions so I'm glad you're
14 heavily involved in this to think about these
15 things.

16 And as you talk I think of different
17 models. So that's the problem. I mean, I think
18 of different models so I think -- I mean, this
19 notion of sampling. I think as we were sort of
20 talking about after lunch, you know, I think an
21 alignment in a lot of ways is overblown. So I
22 might be content at the end of the year to have in

1 a content area three or four really rich
2 performance tests that allowed me to see if the
3 kid could actually integrate and use accumulation
4 of knowledge through the year in a way -- and
5 maybe I do my sampling along the way or maybe not.
6 Or maybe I'm satisfied that there's enough
7 evidence at the end that would allow me to believe
8 that the kid actually knows this.

9 I can think of different models of that.
10 So sampling, you know, I've already knocked
11 comparability and reliability. I might as well
12 kill alignment now, too. Because it does really
13 come down to what are we trying to do here? And
14 if it is trying to further and deepen the
15 understanding that kids develop, I think we think
16 of different models. I think Randy is exactly
17 right. I think about the overarching system and
18 we can think about interchangeable components.

19 Now, it can't be completely like an item
20 bank approach because I think that's totally the
21 wrong approach and nothing makes sense together.
22 But I can think of a lot of ways where you could

1 have certain options. And maybe you want to
2 constrain the options a little bit if the idea of
3 a consortium is that you could have this common
4 system. But I still think you could have options
5 within it that would allow you to still say it's a
6 consortium and it's working together and it's
7 improving the capacity of the states and the
8 schools in that consortium.

9 MR. WISE: When someone -- I forget who
10 it was; it was probably Jeff -- mentioned
11 benchmark as opposed to interim or so on, you
12 know, benchmark in many states is like an
13 accommodation. It's not that it has to be valid
14 for everybody; it just has to be valid for those
15 that need them. And it may be that if schools are
16 doing well as indicated by the summative
17 assessments -- or if states are doing well or
18 entities are doing well, they don't need more
19 continuous information throughout the year to
20 improve their process. Their process already
21 works. But the schools that aren't doing well,
22 they need something more than what they're getting

1 now. And that's where you might want to focus you
2 attention on the different ways of using these
3 interim measures as part of the through-course
4 assessments.

5 MS. WHALEN: I'm just going to ask one
6 more question. So how is that a little different
7 than what we're doing now? So that type of
8 assessment system does exist now and is being
9 implemented now to some good, some bad. I'm just
10 questioning. We're up here talking about
11 innovation. Like, how does that model advance the
12 field and help us get better information?

13 MR. WISE: Well, I think because you're
14 funding it. And, seriously, because it would be
15 built to sort of a higher level of integration or
16 integratability, I frankly don't necessarily
17 subscribe to that. I think the states ought to
18 come to as much common agreement as possible. And
19 if you had a system where everybody used the
20 common pieces throughout the year, it would be a
21 far better system than one where it just focused
22 on improving the schools that really needed

1 improvement.

2 But I think there are different models
3 and states need to be clear in what they're trying
4 to achieve.

5 MS. WHALEN: Purposes.

6 MR. WISE: Yeah. And then how the model
7 that they're proposing will, in fact, likely help
8 them achieve that goal.

9 MR. ABEDI: The idea of consortium of
10 state is not a new idea. We have some
11 understanding, some knowledge of that because of
12 the Title 3 pre-assessments. It's almost three,
13 four years, maybe even more than that that the
14 states actually got together and created a
15 consortia for creating Title 3 assessments. So
16 there are some things -- the good thing about
17 this, when the states want to form a consortia
18 they have some common interest on things like
19 that. So by just forming a consortia that just
20 gives you something. You can start with something
21 that is already there. So we can learn something.
22 We can learn from what we have right now for the

1 Title 3 assessment from these consortia.

2 MR. NELLHAUS: And just a response to
3 the question about what's currently happening with
4 interim assessment. I think it's all over the
5 map. And some of it's very good and some of it's
6 not very good. And so if a consortium were to
7 focus on developing a high quality interim
8 assessment program, you know, we could ensure
9 better quality across the schools using it.
10 Professional development and training can focus on
11 it. We could actually -- states could get
12 involved -- more easily involved in helping
13 schools interpret the data they get from those
14 assessments to improve curriculum and instruction,
15 which is the only reason that you're doing that,
16 as well as trying to get some diagnostic data for
17 individual students to see if they need, you know,
18 additional help in some way. So I think it would
19 just create a much better quality to the benchmark
20 program than we currently have.

21 Also, I think I mentioned in my first
22 presentation here that it's possible that the

1 items that appear on the interim assessments, if
2 they could be calibrated within the summative
3 assessment it would be better interpretation of
4 what the results on individual items mean rather
5 than those items being field tested and developed
6 somehow outside of the system because then it
7 becomes difficult to interpret, you know, what a
8 hard item or an easy item is or how well it's
9 measuring the standard. So on and so forth.

10 MS. WEISS: So -- sorry.

11 MS. WHALEN: I was just going -- one
12 quick question. Do you agree with Laurie that
13 what's preventing us from doing that now is the
14 seed funding for the development of better
15 quality?

16 MR. NELLHAUS: I think it would -- yeah.
17 I think it would help us get to that much more
18 quickly if we had -- and I think a lot of us -- I
19 can't speak for all other states, but I'm sure a
20 lot of states have written this in to their Race
21 to the Top proposals already to basically address
22 this whole issue of interim and formative

1 assessment and how to improve the quality of it.
2 How to link it closer to curriculum. How to link
3 it closer to professional development and training
4 and so on and so forth. So it's, again, now we're
5 talking about a larger system than just an
6 assessment system. Talk about an instructional
7 improvement system. So, you know, we need to
8 think about these things as they relate to other
9 parts of the education process.

10 MS. WEISS: So you guys, I want -- I
11 want to move us on in a minute to the high school
12 thing. But what I'd love to do maybe is just ask
13 Randy -- not just because he's the smartest kid in
14 the class, but also because he happens to have
15 written papers on this that he submitted and we
16 thought were really interesting -- to just maybe
17 take the final word and say if you, based on what
18 we've been talking about for the past few months,
19 but also based on your research and what you
20 believe would work -- if you could waive your
21 magic wand and define the ideal system, what would
22 the pieces and components be just sort of at a

1 high level that you'd think about? And how would
2 you put them together?

3 MR. BENNETT: I guess if I could wave
4 the magic wand, what I would do is create an
5 assessment system that was through-course in
6 nature that had a number of assessments that were
7 given throughout the year. And the idea is that
8 what you would essentially do is create a very
9 long test -- too long to be given at a single
10 sitting -- made up of tasks that modeled for
11 teachers and students the kinds of things that
12 they should be doing in classes. Divide that test
13 into components that you would then distribute
14 across the school year that would, by the nature
15 of the distribution, dictate a sequence of
16 important ideas or core understandings that define
17 that domain.

18 In addition to those tests you might add
19 a project or set of other not -- unstandardized
20 things that would go into a portfolio. And I
21 would take all of those things and do the research
22 needed to figure a way to roll them up into an

1 accountability measure. Note that the more of
2 those things that you have -- through-course
3 assessments plus project and/or portfolio -- the
4 less each counts individually toward that
5 accountability decision. So for any student or
6 teacher or principal or school or district, there
7 is no longer the possibility of being judged by
8 performance on a single day at a single point in
9 time. The judgment of that -- of the
10 effectiveness of those individuals and
11 institutions becomes arguably a fairer one because
12 it takes into account many more pieces of
13 information than the current system that we have.

14 In addition to that summative component,
15 which by its nature now provides interim
16 information because it's distributed across the
17 school year, I would wave the wand further and
18 design a set of model procedures and tasks that
19 teachers could use if they wanted to for formative
20 purposes. And the idea would not be to prescribe
21 what teachers to do (sic) but give them exemplars
22 of what they could do and what they could adapt in

1 whatever ways they thought would work best with
2 their kids in their particular contexts.

3 The idea of the system here becomes very
4 important because what you would ideally like to
5 have is summative assessments that because of
6 their periodicity feed back pointers to teachers
7 about what they should look at more specifically
8 through their classroom formative means or that
9 kids themselves should look at specifically
10 because you want to involve kids in this process,
11 too.

12 And that's about all the wand waving
13 I'll do for now.

14 MS. WEISS: And who -- do you have
15 thoughts on who and how these things would be
16 scored and other parts of the system --
17 professional development or whatever -- just at
18 the high level again that would wraparound it?

19 MR. BENNETT: Yeah, I think you would
20 want to do the scoring of the summative pieces in
21 a way that you could have -- in a way that you
22 could do relatively quickly so that you get

1 information back into the classroom in a way that
2 you could do that would be defensible for
3 accountability purposes. Certainly, if you're
4 going to make decisions about principals and
5 teachers and schools -- if you're going to use it
6 for those purposes you want to be able to do it in
7 a defensible way. So I think you certainly want
8 to do it carefully and effectively.

9 If you could involve teachers in doing
10 that, that would be great. I wouldn't have
11 teachers doing their own classes or their own
12 schools if I were to involve teachers in that
13 component. I would certainly want teachers
14 involved in the formative component. That's
15 critical. That's part of -- and students. And we
16 now have the technology where we could be
17 involving students in doing scoring and
18 understanding rubrics, in practicing how to score
19 other students' productions. Because we want them
20 to be able to recognize what makes for -- what are
21 the characteristics of good work. So I'd
22 certainly want students and teachers to at least

1 have that involvement.

2 MS. WEISS: Well, thank you. That was
3 definitely out there. That was great. I'm sure
4 it got everybody thinking.

5 So now let's move on to our other easy
6 topic. High school.

7 So one of the things that we heard when
8 we had our meetings were a lot of different
9 conflicts about ways to assess and evaluate and
10 hold accountable high schools. One of the things
11 that we know is that the current system that just
12 tests once in high school, as I think Laurie
13 pointed out when he was talking, has been -- to
14 the extent that it can become an actual -- that
15 it's actually correlated with and predictive of a
16 student's college and career readiness, that alone
17 would be a huge improvement over where we are
18 today. But separate from that we had the whole
19 question of how to make sure that the high school
20 curriculum was diverse -- that we weren't so
21 narrowing the curriculum in the high school that
22 we were getting rid of the richness and diversity

1 that I think we want to have out there. But at
2 the same time we had really high levels of rigor,
3 which I think there's a general feeling has fallen
4 off in many high schools around the country. And
5 Algebra I is not Algebra I is not Algebra I when
6 you go to different classrooms, even in one
7 school, much less across different schools in a
8 state. And that the equity issues that arise out
9 of that are just not acceptable.

10 And so in thinking about this it sounded
11 to us as we were talking to people that the
12 end-of-course test was the place to sort of
13 address these questions and these issues. And
14 that perhaps it wasn't an issue of federal
15 accountability. It may be state accountability.
16 It may not be. But that certainly there was some
17 issue of having a consistent high level of rigor
18 across a diverse base of courses and that the
19 end-of-course moment in high school was the place
20 to sort of build systems that helped us address
21 this. And we further had in our heads a couple of
22 different models that we'd heard. One was sort of

1 a top down model. This is maybe not the best way
2 to put it, but almost the Algebra II stuff that's
3 going on is an example of this where everybody
4 sort of came in, came together, and said here's
5 the test that might -- the assessment system that
6 might be the right one to make sure that we were
7 teaching Algebra II to this level. We also though
8 could conceive of the need for diversity prompting
9 a lot of bottom up stuff where it said, you know,
10 I'm teaching this course in Shakespeare and I want
11 to get it certified to this level of rigor so that
12 all the kids taking this know that my test meets
13 this requirement. Or maybe even that some
14 combination of AP tests and tests that were
15 developed by teachers could all be sort of
16 certified at some bar that had some meaning for
17 kids.

18 So we were sort of struggling with, A,
19 is this something that the federal government
20 should play a role in funding, even if it's not
21 part of an accountability system? And the only
22 reason we should do that is if it materially

1 improved the quality of teaching and learning and
2 student outcomes at the high school level. So I
3 guess the question that we have for you is kind of
4 a big one, like is this a good idea? What would
5 be the motivation for high schools, for states to
6 want to do this, for high schools to want to
7 participate in it? If it weren't part of the
8 accountability system are there other motivations
9 or other ways that you could think about this
10 because the last thing I think we'd want is to
11 carve off a substantial piece of this grant only
12 to find out that no one really used it except the
13 early enthusiasts and that it didn't in the end
14 actually serve the goals of improving instruction
15 at the high school level.

16 So that's sort of a big giant mouthful,
17 context- setting piece to help you guys just
18 engage with us around the question of have we been
19 thinking about this right by thinking of sort of a
20 decentralized system that consortia or states
21 could put together with as I think both Jeff and
22 Randy this morning characterized as sort of a body

1 -- a sort of central body that would confer some
2 mantle of rigor on different courses, but that it
3 wouldn't necessarily be something that the federal
4 accountability system required, although states
5 could certainly use it if they wanted to as part
6 of their federal or just state level
7 accountability.

8 So is it worth us doing this? What
9 would be the motivation for using it? If it is
10 worth it, what do we need to do and how do we need
11 to set it up to make it work?

12 MR. NELLHAUS: Just --

13 MS. WEISS: Jump in.

14 MR. NELLHAUS: Kind of a different
15 notion just occurred to me around high schools and
16 why we're asking them to -- why are we looking to
17 develop more tests when through- course assessment
18 is what teachers do every day. Okay, during the
19 course of the year they're giving assessments in
20 chemistry and algebra and geometry all during the
21 year. And then they end up giving students a
22 grade at the end of the year.

1 So the question is, is that grade valid?
2 And what is it valid? What does it really tell
3 us? And the reason we're in this testing business
4 is because we've lost confidence in the grades
5 that teachers are giving.

6 So what if this whole system that we
7 created around high schools really focused on the
8 grades teachers were giving and we were trying to
9 look at the external validity of those grades and
10 develop a system where we could actually somehow
11 certify that those grades have some meaning to
12 them in terms of the students' readiness for the
13 next level of coursework or for college or career?
14 Why are we creating more tests when there are
15 already many, many tests.

16 So that's just a thought, a little
17 out-of-the-box thought about what we're trying to
18 do here. And why don't we get more efficiency and
19 more, you know, effectiveness out of the system
20 that already exists? And how can we accomplish
21 that? Just a thought.

22 MS. WEISS: No, I hear you. And I guess

1 that just to maybe further explain what we were
2 thinking, we were thinking that perhaps those
3 teachers were the very teachers who would submit
4 their things to this body to have the body confer
5 whatever on them. But that we know that -- we
6 know that these kids are coming out not prepared
7 for success, so we don't need to do a lot of
8 studies to know that it's not working today.
9 Right? And the examples we have, like AP that
10 really are organized around a test where that's a
11 standard and an expectation around which
12 instruction is organized, is a lever that works
13 well. So that's why the idea of putting this into
14 a testing system --

15 MR. NELLHAUS: Right. So maybe it's,
16 you know, being able to publish the test and the
17 student work of teachers whose programs are very
18 effective in predicting success at the next level.
19 You know, that's what's going to matter. Is
20 knowing what are the questions that these teachers
21 are asking their students every week? What is the
22 work that they require? The quality of work that

1 they require that gives a student an A or a B or a
2 C.

3 MS. WEISS: And so if this body -- if
4 that was the evidence that a teacher submitted and
5 this body conferred whatever on it and there was a
6 research plan that went along with it and then
7 those things were available to any teachers in the
8 consortium or outside the consortium to use, does
9 that like -- so what are the characteristics that
10 might get us to the place where we'd have a system
11 for doing it instead of islands of excellence that
12 are posted on the web?

13 MR. WISE: Well, if you had an
14 independent body that was certifying these tests
15 as being appropriate or not, you'd still have to
16 have some common criteria that everyone agreed to
17 that the body would use to decide it was or it
18 wasn't. So it seems like the really hard work and
19 heavy lifting is to get across states. And the
20 real equity issue is to get states to agree on at
21 the end of an Algebra course what should students
22 know and be able to do. And once you've done

1 that, whether you have, you know, building a good
2 test to that probably is -- has a greater impact
3 than allowing a lot of tests to bubble up. But
4 you at least have the criteria that you would want
5 to use in judging the bubble up tests. I'm not
6 the bubble -- I mean, the locally developed tests
7 -- as to their adequacy. So that you do need to
8 do this heavy lifting first I think of the
9 content. Agreement on the content objectives.

10 MR. MARION: You know, we -- back in the
11 days when I was in Wyoming we actually did
12 something like this. It was the Wyoming,
13 actually, graduation system which is still
14 (inaudible) still in existence. It's called the
15 Body of Evidence System. It's where there's not a
16 common state exam. Each district has to certify
17 that their students have met important graduation
18 requirements. And this is going on in Rhode
19 Island, as well. Wyoming has just been at it a
20 little bit longer.

21 And one of the things, Jeff, as you
22 mentioned, this sort of certifying courses, when I

1 was integral in the design of it, we said these
2 are the standards that kids need to meet. And
3 we're going to then judge the set of assessments
4 by a variety of technical criteria that people go
5 through. But we said, you know, you could use
6 completely sort of common standardized approach.
7 You could use course space approaches. In any
8 case, one of the things that we did find out, we
9 ended up creating this huge project to invent
10 these -- they really were curriculum-embedded
11 tasks because the current assessments in the
12 classes were actually quite lousy. And so nothing
13 was going to bubble up, and when it did the
14 bubbles burst fast. So we ended up creating this
15 huge task bank.

16 Importantly, it didn't go to some -- and
17 so our neighboring state Nebraska had a similar
18 idea and they actually had their district
19 assessments evaluated by a body convened by
20 Burroughs at University of Nebraska. In Wyoming
21 we did it very differently. We had sort of a peer
22 review process. Actually, a little bit modeled

1 off the ISA peer review process. But we had teams
2 of district and school folks come together once a
3 year initially or twice a year initially. And
4 then it -- actually only as districts needed to be
5 reviewed. But that also was a great way to
6 internalize the requirements and what does it take
7 to develop a good assessment system. Or in the
8 case that Laurie was talking about, what does it
9 take to develop appropriate criteria for Algebra I
10 or Biology I?

11 So there's models out there for this.
12 And I really appreciate the way that you're
13 talking about this -- that it's not this common
14 assessment for end-of-course assessment for
15 everybody in the consortium, but it could be one
16 of many within the consortium that meet this sort
17 of threshold of rigor, which I think is what we
18 all want to get it.

19 In terms of the use thing, this is again
20 where you're not just throwing the money out
21 there. Your people have to apply for the money.
22 So as part of the application I would say why do

1 you want to do this? How does this set of
2 assessment or this approach that you're taking for
3 high school assessment fit with your theory of
4 action? And then how will the way that you design
5 these assessments and design the implementation of
6 them -- that means whether it's part of the
7 accountability system; whether student level,
8 school level, teacher; whatever it might be -- how
9 is that going to fit with getting kids engaged and
10 improving the quality of the course, et cetera,
11 like that? I think people could specify that and
12 I think that would be a useful exercise.

13 So I, actually, if I was doing it -- and
14 you keep talking about this as not being part of
15 the accountability system -- I think states might
16 want to know that it would be allowed to be part
17 of the high school accountability system so
18 they're not doing all this stuff and then they
19 have -- now they have this 11th grade, you know,
20 NCLB test in math and reading and science. They
21 would probably want to know at least that this was
22 allowed to be considered as part of the federal

1 accountability system as far as you could see on
2 that. So I think there's a lot of merit to this
3 approach. And I think people could propose really
4 smart things.

5 The last thing I'll say about this is
6 we've done high school assessment and
7 accountability really badly in this country.
8 We've not take account of the uniqueness of high
9 schools compared to 3 through 8. And this would
10 be a way to at least allow us to move a little
11 forward in that area.

12 MS. DeSTEFANO: One of the areas in my
13 own state that I think this would be attractive is
14 the issue of kind of making high school meaningful
15 and dual enrollment opportunities for students.
16 And I could see this mechanism as one way to
17 identify courses that were potential dual
18 enrollment candidates. And also to address the
19 idea of the amount of remediation that has to
20 occur when students enter community colleges or
21 even four-year universities. So to use these sort
22 of certified courses or certified programs of

1 study as one way of holding high schools
2 accountable for the exit characteristics of
3 students in these courses.

4 MR. ABEDI: I think a missing component
5 here is not having enough information on the
6 current system. I think that would be extremely
7 helpful to see exactly what are the issues that
8 the current system that we have at the high school
9 level and what are the areas that can need
10 improvement. So rather than focusing on creating
11 the assessment or (inaudible) consortia for making
12 assessment, point out exactly more -- focus on
13 data collection and finding out exactly what are
14 the issues -- what are the issues that need to be
15 fixed.

16 MS. WEISS: I mean, Jeff, what do you
17 think? Is there a motivation for high schools --
18 I mean, for states wanting to do this or for high
19 schools wanting to adopt it? Or how should we ask
20 questions if we decide to do this that would
21 create the right motivation maybe is a better way
22 to ask it?

1 MR. NELLHAUS: I think there definitely
2 -- there definitely has to be a motivation if it's
3 not going to be for accountability uses per se.
4 And there is student motivation so there are
5 states that have high school exit exams. There
6 are states who might use these courses to qualify
7 students for dual enrollment. There might be
8 states who want to use these assessments to
9 qualify students for certain scholarships,
10 opportunities, and so on and so forth.

11 But there may also be just if it's low
12 stakes schools that just want to do this to
13 improve the quality of their programs. So I
14 think, you know, you may be able to get buy-in
15 from schools as long as they can see it's a way to
16 get professional development training and to build
17 quality.

18 MS. DeSTEFANO: I think it could have
19 big curricular impact on certain courses -- high
20 stakes courses, like Algebra, like some of the
21 Science courses. So I think there is a gain here.

22 SPEAKER: I agree.

1 MS. WEISS: Go ahead.

2 MS. WHALEN: Can you guys see any
3 advantage to having this be a system where by
4 having multiple states come together as part of
5 the development of this system, trying to solve
6 this problem at a secondary level versus kind of
7 what, Scott, you described as a one state
8 solution? And if so, can you guys talk about
9 maybe some of the ways that can help play out and
10 advantages to that?

11 MR. MARION: I'll just offer the reason
12 why it was a one-state solution is we didn't know
13 from consortia in those days, so it was -- and
14 this was novel then. It was just going across
15 districts. That was, you know, to actually build
16 this common task bank across districts was -- and
17 it was a small state. It still is a small state.
18 Its capacity was a challenge to do that. So I
19 think it's a huge bang for your buck in working
20 across states. And, again, if I was going to do
21 something like this, this is a case where I would
22 involve teachers. So there's a lot of talk in the

1 notice and all the other things about teachers
2 involved in scoring. And it leaves a lot to be
3 desired. I think you get something for it. In
4 this development part I think you get a lot. To
5 think about units of study that go along with
6 these tests. So I think there's a lot to be
7 gained from working across consortiums. It
8 doesn't mean that if you have 20 states in a
9 consortium everybody has to be equally represented
10 at each of these content meetings, but you can
11 sort of figure out which -- how to structure that
12 organization and governance. I think you could
13 gain a lot.

14 MR. WISE: And I see a couple, three
15 other reasons why states coming together could do
16 a better job than by themselves. I mean, there's
17 the obvious one -- pooling resources for
18 developing, maintaining higher quality systems.
19 But another one is it lets the state off the hook
20 a little bit for being totally responsible for
21 defining the content. So in the kneecap
22 consortium, you know, Peter McWalters from Rhode

1 Island has all the time said if the state develops
2 the content requirements for a course, then they
3 can be accused of biasing it up or biasing it down
4 to look good or so on. But if it's an outside
5 group, a consortium that's developing these, then
6 clearly the state is not doing this for any self-
7 serving gain purposes. And it does, in fact, let
8 them off the hook a little bit. So that's another
9 reason besides just the common reason. And then,
10 of course, the equity issue. We want the same
11 expectations for students across states. And if
12 the states don't go together, we're not getting
13 very far with that.

14 MS. DeSTEFANO: Just another comment.
15 This is definitely a place where higher education
16 and workforce representatives need to be very
17 engaged with defining the standards and maybe even
18 doing some of the review and vetting of the
19 instruments.

20 MS. WHALEN: Just -- I'm sorry. Go
21 ahead.

22 MR. BENNETT: This is an instance, I

1 think, that really would benefit from multiple
2 consortia trying different approaches. Because
3 take the two limit cases. So one limit case is
4 that a consortia has a set of end-of-course tests,
5 but it's the same test for any given course to use
6 throughout the consortia. The other limit case is
7 the Let Freedom Ring case and that, you know, any
8 teacher can propose any assessment that is
9 certified by the external body and approved by the
10 external body for use.

11 So then the next question is, well, what
12 happens? What's the impact of doing that in those
13 two different cases? And clearly I think the best
14 way of looking at that impact is going to be
15 through a set of very intensive case studies to
16 find out what's going on. But one could imagine
17 that they would have very different effects. And
18 some of those effects may be good, and some of
19 those effects may not be good. So you can imagine
20 in the case of the Let Freedom Ring case that
21 employers might have some difficulty in knowing
22 what a credential meant when it was based on this

1 vast multiplicity of possible assessment
2 mechanisms.

3 On the other hand you might argue that
4 in the case where there was this vast multiplicity
5 there was much more freedom to innovate among the
6 teachers and perhaps better development as a
7 result. And perhaps higher achievement as a
8 result. But, you know, that's just a possibility.

9 MS. WEISS: Two hypotheses. Right.

10 MR. BENNETT: Yeah. Yeah.

11 MS. WHALEN: So can I ask just a quick
12 follow-up question? We've been talking about this
13 as an end-of-course system. Could you guys
14 envision this in any way where it could be
15 components of an assessment system for each course
16 or each subject area that's administered at
17 different times? Or would you think that there's
18 value in just keeping it as an end-of-course
19 system for high school courses?

20 MR. NELLHAUS: In modules?

21 MR. WISE: I think the reason it came up
22 as end-of- course is in contrast to some of the

1 state assessments which are given with no
2 relationship to a particular course.

3 MS. WHALEN: Right.

4 MR. NELLHAUS: End-of-grade.

5 MR. WISE: But for all the reasons we
6 thought through-course assessment might be good,
7 there's no reason that it shouldn't also apply to
8 high school courses as well.

9 MR. NELLHAUS: Right.

10 MR. BENNETT: I imagine that you have
11 your own action theory that there's one implicit
12 in why you're thinking about a more standardized
13 approach at the three through eight level toward
14 accountability and a much more open one at the
15 high school level. So I'm imagining that you have
16 -- you believe that there are impacts of doing
17 each of those that are best served by those
18 approaches or else there's a political reason.
19 But if there is -- an action political, yeah?

20 MS. WEISS: Right.

21 MR. BENNETT: But if there is an action
22 theory, it might be helpful -- theory of action --

1 to make it more explicit in the notice to better
2 contextualize why the approach you've chosen for
3 three through eight is significantly different
4 from the approach you're choosing for high school.
5 And that will be, I think, helpful to the
6 consortia in responding and creating their own
7 theories of action for how to get to the effects
8 that you're suggesting ought to be achieved.

9 MS. WEISS: So, yeah. I think we heard
10 you that we need to put forth our theory of action
11 just like we need to ask states and consortia to
12 tell us theirs. So we will figure out a way to do
13 that.

14 Just to clarify, I do think that what
15 we're thinking of in the larger comprehensive
16 whatever it is -- through-course or other
17 summative assessment system -- we're not just
18 thinking three through eight. We are thinking
19 three through college and career-read and then
20 there's this other question of high school
21 instructional improvement at-large that we're
22 trying to also enable. But, yes, we will try to

1 make this -- we will write it down.

2 Other questions about this? Okay. So
3 another issue that we wanted to come back to was
4 this question of additional research. And one of
5 the questions that we had since you have already
6 said that John Easton could put in the IES budget
7 all of this new research work is aside from who is
8 paying for that, the question of what should
9 happen where? So what are the kinds of things
10 that we should require states and consortia to do
11 as part of their applications integrated into the
12 assessment systems that they're proposing? And
13 what by rights should happen sort of sitting next
14 to it on the side and somewhat externally and why?
15 How do we think about what the difference is
16 between those two and what motivates you to put
17 something in one category or another?

18 Go ahead.

19 MR. WISE: You guys ask good questions.

20 MR. EASTON: May I complicate it and ask
21 another version of that question?

22 MS. WEISS: Absolutely.

1 MR. EASTON: So I was really listening
2 to all of your perspectives on the research
3 development evaluation roles. And it really is
4 kind of neat because Laurie and Jeff talked a lot
5 about either a process improvement or research and
6 development built into the work. And then Randy
7 and Scott talked a lot about validation studies
8 and kind of construct validation at-large. And
9 then Lizanne talked about this meta research going
10 across. And then Jamal brought in the quality
11 control aspect.

12 So I think that's -- I'm re-asking
13 Joanne's question because I thought this really
14 did make a nice framework. And I wonder if you
15 can respond to that and Joanne's question at the
16 same time. What belongs inside? What belongs
17 outside? And how do we accomplish all this work?

18 MR. ABEDI: I think one thing that's
19 important to keep in mind is that as I mentioned
20 in our brief discussion before this -- the outcome
21 of research is not going to come out on time to
22 actually have much impact on this -- on the

1 consortia award. So there are some major
2 components -- research components that need to be
3 done during or prior to the consortia award that I
4 think should be part of the application. For
5 instance, the research on the -- what will be the
6 best way to create a composite in the components
7 or things like that. Or how to identify what are
8 the redundancy -- specific contributions through
9 all of these. These are something that has to be
10 part of the application. Even some of those
11 generalizability things you mentioned, those are
12 essential in order to actually finish the work or
13 development. But there are some long-term issues
14 that need to be addressed by research and those
15 are the issues that need to be (inaudible).

16 Again, it's not going to have much
17 impact on the current development, but it's going
18 to be extremely useful for future. Hopefully, we
19 are going to have more future activities
20 (inaudible) and plans for this type of work.

21 MR. WISE: So one way of thinking about
22 what should be part of the grants and done by the

1 people doing the grants is sort of who could do it
2 best. Many State Departments of Ed and states
3 don't have huge a track record of funding and
4 managing research. But John, you've got these
5 centers that are pretty good at it. And so you
6 can ask is there an area of research where it
7 would benefit not just one consortium, but several
8 consortia. And would it be more effective to do
9 that once and well through a center versus having
10 to embed it in the work of each of the individual
11 consortia themselves.

12 A second thing that's really a critical
13 thing is that it's not just research for research;
14 it's research that gets used. And so you may want
15 to consider building into the mission of the
16 centers or the mission of whoever is doing the
17 research not just to answer the question, but to
18 work closely with the consortia and the states and
19 get graded almost. But sort of a part of their
20 mission is to see that the outcome of the research
21 actually gets used.

22 MS. WEISS: Other thoughts?

1 MR. MARION: Yeah. I mean, I do think
2 that John's classification scheme was accurate. I
3 think that as part of the consortium proposals,
4 like Jeff said, unless it's required it'll get
5 spent on other things. So I would require a
6 validity plan -- a validity evaluation plan as
7 part of any consortium proposal and dollars set
8 aside to actually carry out proposed studies
9 related to the various aspects of the theory of
10 action or components of the system. And those
11 would serve important purposes. It would
12 certainly help us learn a lot about how this thing
13 worked and the threats to validity, et cetera, et
14 cetera. But also as Lizanne talked about this
15 notion of sort of the formative evaluation
16 component, you wouldn't wait until the end of the
17 four years to say you've got an invalid system.
18 Too bad. But, you know, along the way you get
19 this feedback, course correction kind of stuff.

20 I think -- trying to think about the
21 right meta issues, if you will, that would be most
22 beneficial -- those that would be, I think, that

1 would involve all consortia. Things like -- I'm
2 just throwing things out, things like equating.
3 All right? How to best, you know, incorporate
4 these innovative designs and equating designs
5 around these innovative item types and innovative
6 assessment types. I think -- and I don't know how
7 you would do this kind of study, but I think we
8 need more study of accountability systems and how
9 they actually interact with and affect the
10 assessment systems and the learning systems. And
11 which ones are working or not. Now it's hard to
12 set up the right kind of experiment to do that,
13 but there's other ways -- I'm thinking about
14 studies -- there's comparability; there's
15 performance assessment issues; there's a range of
16 issues that are going to cut across any
17 computer-adaptive testing, that are going to cut
18 cross any of the consortium. And I think those,
19 you know, make sense to rise up to this meta
20 level. It's not to say that they won't be done
21 within each consortium.

22 So if I'm doing a consortium, part of my

1 validity evaluation better say something about
2 equating studies if I have to compare you to your
3 scores. All right? But I'm looking within a
4 specific context so the advantage of a center
5 could look across different types of
6 implementation and sort of perhaps get more
7 generalizable knowledge about this. So I think
8 your classification scheme is the right way to
9 think about this, so.

10 MS. DeSTEFANO: And I think it's very
11 important for there to be a percent allocation or
12 a certain amount of money that's set aside for R&D
13 at the state level for all the reasons that people
14 talk about. And a portion of that needs to be an
15 agreement to work with these meta centers because
16 it could be that the consortia will be asked to
17 provide data to serve on kind of expert review or
18 other kind of panels. And so I think that synergy
19 between what's happening in the state and what's
20 happening in the center is a really important
21 piece.

22 MR. BENNETT: It would be good to the

1 extent possible if you could coordinate your two
2 funding programs so that they could be mutually
3 supportive of one another in the following sort of
4 way. So, for example, you want to have centers
5 doing meta research, but it's silly to have them
6 do meta research on ideas that weren't funded or
7 it would be great if they could do meta research
8 on ideas that were funded as part of the main
9 competition where those ideas are important ones
10 deserving a meta research. So, for example,
11 there's a whole list of issues around
12 through-course summative assessment. It would be
13 great if there was a consortium that was pursuing
14 that model and if there was a meta center that was
15 focusing on addressing some of those more
16 difficult issues and problems and sharing data.
17 And then the meta center in turn concentrating on
18 helping to solve some of the both near term and
19 longer term issues for that consortium.

20 MR. WISE: So one way you could
21 facilitate this would be to sort of write in and
22 plan annual or semiannual conferences where you

1 brought people from the consortia and people from
2 the centers together and allowed some time for the
3 consortia to talk about the problems they're
4 having and allowed some time for the people in the
5 centers to talk about the answers that they're
6 getting so far. And to listen to the problem so
7 that their research agenda can pick up on those to
8 the extent possible.

9 MS. DeSTEFANO: I think another value
10 added of the meta centers is dissemination. It's
11 hard for states to keep all the balls rolling and
12 publish and get out white papers and disseminate
13 models. So dissemination should also be a
14 function of the meta center.

15 MR. ABEDI: I can think of two different
16 types of consortia. One research consortia and
17 one development consortia because various states
18 may have the same questions, same research needs,
19 that they can get together and form a research
20 consortia in addition to the consortia that we are
21 talking about. So that's one point I wanted to
22 make.

1 Another one is that we need not to
2 forget about the value of existing data. The
3 existing data actually could be extremely helpful
4 of what the states have done, what we can learn
5 from it, and so forth. So we don't need to
6 repeat. Some of these existing data are going to
7 be valuable in order to tell us more about what we
8 are looking for.

9 MR. WISE: Some of the existing data
10 provide a baseline for improvements to
11 accountability systems.

12 MR. ABEDI: Exactly.

13 MR. WISE: If you don't know how bad it
14 was to start, it's hard to know how much better
15 it's getting.

16 MR. BENNETT: It might be worth thinking
17 about what the best way to coordinate innovation
18 funding is. So, for example, you're going to make
19 grants to consortia to put into operational
20 practice a next generation assessment system in
21 four years. But you also want to have them or
22 someone working on what happens in year 5, 6, 7,

1 8, 9, 10. It would be great if one of the centers
2 -- one or more of the centers -- could be working
3 in partnership with one or more of the funded
4 consortia to begin to, you know, put that
5 innovation program into place so that there is a
6 pipeline beginning that will feed into that
7 consortium at a later date when it's ready to take
8 on that new innovation.

9 MR. MARION: So it sounds like when you
10 get a bunch of researchers together asking about
11 research needs there's always a lot.

12 MS. WEISS: I know the set aside you
13 were going to propose was 95 percent, something
14 like that?

15 MR. MARION: Yeah.

16 MS. DeSTEFANO: 99.5.

17 MR. MARION: We joke. I do a lot of
18 evaluation. Evaluators always ask for 10 percent
19 and they're thrilled if they can get 5 percent.
20 And usually it's not quite enough.

21 But in this case, Ann, you asked this
22 question earlier about how to make sure we don't

1 paint ourselves in a corner. So this is -- at
2 your level really, at the U.S. Department level,
3 how to make sure that after four years we're in a
4 better position than we are now. And that wasn't
5 really -- I mean, peer review is part of NCLB, but
6 not -- the whole validity stuff was ignored for
7 the first four or five years. States were running
8 as fast as they could and it was up to them to do
9 it. There was no set aside. So I think if you
10 build this in, I think almost no matter what
11 happens with the consortia, we'll be in a better
12 place four years from now than we are now in terms
13 of our knowledge.

14 MS. WEISS: Do you have --

15 MS. WHALEN: Sorry. The one other
16 question on top of Joanne and John that I just
17 wanted to ask because I realize this may be a
18 loaded question asking with a bunch of researchers
19 here, but knowing that we do have a limited amount
20 of funds, how would you prioritize everything you
21 just said?

22 MR. MARION: Well, I would put a high

1 priority on validity arguments, evaluations of
2 theories of action as part of the funded consortia
3 first. And I don't think that that -- considering
4 the amount of money that is potentially there to
5 fund consortium and who knows what's happening
6 with this new announcement, I think that that is
7 money well -- necessary money. Especially if we
8 want to do innovation. We're not that good at
9 evaluating the same old, same old. But if we're
10 going to do innovation we better really be
11 thinking about evaluating and constructing these
12 validity arguments the way that Randy, you know,
13 really eloquently laid out. That it would allow
14 people to evaluate the particular claims as part
15 of this system.

16 It doesn't have to be the end of the
17 world, and I think -- I mean, I don't know,
18 listen, I would say, you know, in the area of less
19 than 5 percent of the budget -- of a consortium
20 budget. You could do a lot. And maybe even less
21 than that. HumRRO has got a ton of experience in
22 this kind of --

1 MR. WISE: In doing things cheaply?

2 MR. MARION: And well.

3 MR. WISE: One of the lessons I've
4 learned over the last couple of years is the value
5 of a balance portfolio. So just because one thing
6 seems important doesn't mean all the money or
7 funds should go there. But you need to figure out
8 how to strike a balance between -- and I would
9 certainly agree that the validity argument
10 research is critical. But you also need some
11 research on longer term problems that's going to
12 support, you know, and build the infrastructure to
13 support continued improvements after the four-year
14 grant.

15 MS. DeSTEFANO: And think of the
16 consortia as your portfolio. You know, hopefully
17 they're going to gravitate to areas where they
18 have interest, expertise, some bit of a track
19 record. And so you can look across them and sort
20 of see them as your portfolio. What looks
21 promising? What looks like investment? So I
22 wouldn't narrow it in the RFP now. I would see

1 sort of what arises from the consortia.

2 MS. WEISS: So let me pile on these
3 questions with one more permeation of this. Is
4 there anything that we could or should do in the
5 notice that would enable the research community
6 at-large, without our funding, to be able to get
7 into the action and start studying everything from
8 learning progressions to individual student
9 patterns of effectiveness through all of this
10 data? Like, should we say that the student data
11 needs to be depersonalized? I mean, do you have
12 words for us that we should put in here that would
13 make this data as it's going to be collected
14 available to the research community for study?
15 And if that is a good idea -- well, first of all,
16 is that a good idea? Will research -- will the
17 research community come, take it apart, look at
18 it, and start feeding what it learns back into the
19 practitioner community? And what do we need to
20 say about it to enable that to happen?

21 MR. WISE: You're asking a bunch of
22 researchers should you make data available for

1 research?

2 I think you might want to look at it as
3 a condition of the grant that there is some system
4 for de-identifying individual student data and
5 making it available -- readily available. It
6 takes a little bit of effort. Usually not a huge
7 amount. But make the data -- relevant datasets
8 available for research and encourage the states in
9 the consortia to promote the use of those data for
10 research.

11 MR. ABEDI: If you already have
12 experience using confidential data from NCES,
13 NAEP, for instance, we can have access to that
14 under (inaudible)? So actually, these deal with
15 confidentiality issues, so we know how to deal
16 with those type of issues and how to make data
17 accessible to researchers without jeopardizing the
18 confidentiality.

19 MS. WEISS: Right. Right. It's
20 sometimes easier if you specify it at the front
21 end than if you try to build it on at the back end
22 so I'm just trying to make sure we know what the

1 right requirements or questions are.

2 MR. WISE: Yeah.

3 MS. DeSTEFANO: I think that language
4 would certainly be useful in the consortium
5 applications, but certainly in the center
6 applications. One of those functions should be to
7 produce datasets that the general public can have
8 access to, a la NAEP or NCES.

9 MR. NELLHAUS: That said, we are limited
10 by FERPA.

11 MS. WEISS: Right. We would have to --

12 MR. NELLHAUS: So, you know, you can't
13 --

14 MS. WEISS: Right.

15 MR. NELLHAUS: -- ask us to do anything
16 that we can't do under federal guidelines.

17 MS. WEISS: Yeah. Yeah.

18 MS. WHALEN: Can I ask -- I apologize.
19 I don't remember who said this, but at some point
20 someone said, today, have all of the applicants or
21 grantees -- it wasn't distinguished -- stand up
22 and share their theory of action and how the

1 assessments fit into that theory of action -- to
2 have that kind of poked and prodded and improved.
3 When do you see the timing of that playing in?

4 MR. BENNETT: Yeah. As early as
5 possible because the idea is to expose that theory
6 of action broadly so that it can be evaluated and
7 taken apart. And if there are indefensible ideas
8 in it, then that's the time. You know, the
9 earlier the better. And give that consortium an
10 opportunity to rethink the theory, rethink the
11 components. And, you know, to the extent
12 possible, redesign. What you would like is for
13 bad ideas to be killed early on and good ideas to
14 be identified, singled out, and vigorously
15 pursued. So the sooner you do it the better.

16 MR. WISE: I would add as often as
17 possible on top of that. Because there are some
18 parts you can just look at the theory and you can
19 test it without too much data and then other -- as
20 data continue to be collected as part of the
21 development process there are more things that you
22 can continue to --

1 MS. DeSTEFANO: So ideally prior to the
2 application.

3 MR. NELLHAUS: So, I mean, this is
4 interesting. So in order to be awarded the
5 application it might be that you don't have to
6 have it entirely right. That you're going to go
7 through a process in the first several months
8 especially after you're awarded it to refine it
9 based on feedback you get from these open
10 sessions. So that might be a way to address that.

11 MR. WISE: And I would, in fact, be real
12 suspicious of an applicant who said they had it
13 entirely right.

14 MR. NELLHAUS: Right.

15 MR. MARION: But would you use it as
16 part of the evaluation criteria for the
17 applications, at least a preliminary theory of
18 action? Or perhaps how you're going to go about
19 designing a theory.

20 MR. WISE: Both the preliminary theory,
21 but also the process for continuing to improve it.

22 MS. WEISS: And I think we've heard from

1 a number of our things, like, especially the
2 Consortium Management Group, that we have to ask
3 for very, very focused and specific answers to
4 detailed questions in this. It's not that they
5 have to necessarily -- the end product may look
6 different, but having a sort of theory of change
7 and design that matches it with some nontrivial
8 level of specificity is, I think, the direction
9 that we've been prompted by everyone to go,
10 knowing that over time as we learn things that
11 should be allowed to change in response to what
12 we've learned, but not that we delay the
13 specificity till later.

14 MR. MARION: Could I ask you guys a
15 question? I'll ask it. You can choose not to
16 answer.

17 What do you envision as the end product?
18 Because that might help someone. At the end of
19 the four years is it a full-scale implementation
20 of an assessment system? Or is it a design for an
21 assessment system that has some pilot testing and
22 things like that? Or we could wait two months and

1 find out.

2 MS. WEISS: No, no, no. I mean, no, we
3 believe that at the end of the four years we need
4 to have a system that is ready for wide, you know,
5 wide-scale implementation. So it's been through
6 field testing and it's ready to be deployed.
7 Yeah.

8 MS. DeSTEFANO: And certainly with the
9 reauthorization of ESEA there would be reason to
10 expect that there would be a serious
11 reconsideration of the theory of action in light
12 of that.

13 MS. WEISS: Yes, right. It's possible
14 that that would prompt one of these meetings that
15 you just described to occur.

16 MS. WHALEN: So I just wanted to
17 revisit, Joanne, your question about how we make
18 some of this data publicly available for
19 researchers. I'm just wondering, Jeff, you
20 mentioned the FERPA issues from a state's
21 perspective since you are going to be the one on
22 the other side of this producing this data to be

1 consumed. Do you have any other thoughts, ideas,
2 questions, concerns about this?

3 MR. NELLHAUS: No, I mean, we've made
4 data pretty available to researchers. But, you
5 know, we -- I don't know. You just can't make --
6 we post what we call research files on our website
7 that anyone can use. But the way we do it is we
8 de-identify it by not having the student's ID.
9 But we have one set where school and district data
10 is identified, but not demographic. And in
11 another situation when the demographic information
12 about the student is indicated, but the school or
13 district isn't. And many researchers want both.
14 And that's when we get into trickier situations.
15 When you can link the school with the student's
16 race and gender --

17 MS. DeSTEFANO: Have a small end.

18 MR. NELLHAUS: Yeah. You get a very
19 small end. So in that case, you know, we have to
20 get into more detailed agreements with the
21 researchers and actually, they're agreeing to
22 answer some research questions that we see as a

1 high priority. We're just not giving out the data
2 to anyone at that point.

3 MR. MARION: Yeah. In defense of
4 states, I mean, they're busy. Even though you can
5 post stuff you're still going to get the requests.
6 And so who's -- anybody that puts out a shingle is
7 a researcher and they should have access. So it
8 becomes tricky. And then who controls what? This
9 is going to be an interesting issue. So I'm all
10 for having data available for vigorous programs of
11 research. I think that should be part of the
12 consortium proposals to discuss how they're going
13 to do that. How they're going to handle requests
14 for data. You know, who's going to be responsible
15 for that? And will they have a vetting board?
16 Will they have a mini-IRB within their --
17 something like that because they can't just be
18 giving out the full data file, like Jeff just
19 said, that lots of people would want to just
20 anybody.

21 MR. NELLHAUS: Interestingly, most of
22 the requests we get for data is for program

1 research, not to research the assessment system
2 itself. So it's going to raise some different
3 questions. And, you know, for some states and
4 consortia that could be threatening, you know, to
5 -- we're going to evaluate your assessment
6 program, not the program that the schools are
7 implementing to see if this particular curriculum
8 works or so on and so forth. It's a different
9 take on this.

10 MR. EASTON: They're going to want item
11 strings and they're going to want secure items.

12 MS. WEISS: Well, so in our last minute
13 any final words? And you're not allowed to say
14 good luck and better you than me because Ed
15 (inaudible) has already used those up in their
16 leads and closings.

17 MR. ABEDI: I just wanted to mention
18 something. We talked about assessment issues and
19 (inaudible) factions and other things as applied
20 to everyone, but there are some issues concerning
21 a specific subgroup of students. I know we have
22 discussed that in the previous discussion or

1 something like that. But these issues need to be
2 -- for instance, I just want to give you an
3 example. Even though these -- I mean, through the
4 course, different components provide much better
5 opportunity for ELL students and students with
6 disabilities to present what they know and can do.
7 But there are some limitations, some issues that
8 multiple assessment can create. So are these
9 issues something that you are thinking about or do
10 you think discussion today gives you sufficient --

11 MS. WEISS: Actually, so, those are
12 definitely issues we're thinking about. And I
13 think our feeling was that we got such good input
14 when we did the ELL and SWD panels that we didn't
15 walk away with a lot of questions about those the
16 way we did about this. So the fact that we didn't
17 put in questions about that is not an indication
18 at all that we have deemphasized it or don't think
19 it's important. It's that we actually walked away
20 feeling like we learned a lot from those panels
21 and had most of the information that we needed in
22 order to write the notice. And these were the

1 questions that we were less certain about.

2 MR. ABEDI: So, for instance, my
3 question is in theory of action do we need to pay
4 attention to this or just do something that
5 applies to everyone? Just one example.

6 MR. WISE: So several things I've heard
7 here that sort of reinforce what I think you may
8 already know, but you need to be real clear on the
9 goals that you hope will be achieved through these
10 grants. And then be a little open to innovation
11 on the methods that the states propose, but
12 require a clear theory of action for how what
13 they're proposing will achieve those goals and
14 require both early and often validity research to
15 check and demonstrate that theory of action. And
16 then finally, have some mechanism for midcourse
17 corrections as the validity research comes in and
18 isn't precisely what was hoped for or predicted
19 initially.

20 MR. NELLHAUS: That's great. Perfect.

21 MR. MARION: Yeah, what Laurie said.
22 But just one little tweak is that I would -- so

1 even though we've given you the assignment of
2 developing the Department's theory of action, I'm
3 not sure that you could even do it in the month
4 probably you have to get this out. But if you
5 could get exceptionally clear about the goals and
6 purposes, that would be a huge help to the
7 proposal writers. And then you might want to play
8 out theories of action, you know, on your scrap
9 paper to see if, in fact, this actually might --
10 is there any way that you could see that this
11 could be doable what you're asking states to do.
12 So, but it's important, I think, getting really
13 clear about the goals and purposes.

14 MR. NELLHAUS: I don't want to be
15 redundant. They said it.

16 MS. WEISS: Great. Well, thank you all
17 very, very much. I'd love to give you a big round
18 of applause. That was amazing today.

19 So on behalf of the Department, I thank
20 you all and all the people who we have spoken with
21 over the course of this and received input from
22 have just done an amazing job of giving us a

1 fabulous education that hopefully you will see
2 reflected in the final notice that we put out, as
3 well as in the proposals and applications that come
4 in from these consortia. So thank you very much.
5 Thanks to all the states for hanging in there with
6 us through all of these. You're now dismissed.

7 We're going to take a quick break.

8 We'll be back here -- let's come back a couple of
9 minutes before 4:00. We have a full slate of
10 public speakers who are going to be with us so we
11 just need --

12 (Recess)

13 MS. WEISS: Thank you. So let me just
14 give quick instructions to our speakers so that
15 everybody knows how this will work.

16 We've given everybody ahead of time --
17 told them that they would have five minutes to
18 speak. On the podium you will see a countdown
19 timer that will turn from green to yellow when you
20 have two minutes to go. And then turn to red when
21 you're out of time. And I think everybody has got
22 a number and is coming up in order.

1 Okay, so this is just a separate --
2 okay. Okay, yes.

3 MS. IRISH: I didn't realize I was
4 supposed to get there already.

5 MS. WEISS: I didn't either, so that
6 makes two of us.

7 MS. IRISH: That's the problem with
8 being the leadoff hitter.

9 MS. WEISS: That's right.

10 MS. IRISH: My name is Myrtice Irish. I
11 am a retired teacher of the Los Angeles Unified
12 School District. Actually, I'm a retired central
13 office administrator. I hate to admit that. And
14 I still serve as a representative of the Community
15 Advisory Council for that district. And I'm also
16 an assistant professor out in California
17 University preparing special ed teacher
18 candidates.

19 And I tell you this date -- 1963 is the
20 day I started and stepped into my first classroom.
21 And I tell you this because I have an extremely
22 long view of the unanticipated particularly

1 consequences of trends in public education. The
2 anticipated were well published; the
3 unanticipated, of course, were dire results. And
4 I also want to tell you because of that long and
5 varied experience that without exception, learning
6 only takes place in the relationship developed
7 between a teacher and their students in a
8 classroom as they pursue the concepts and the
9 skills of the curriculum. It occurs nowhere else.
10 And I am terribly afraid of the emphasis on
11 teacher uniformity that is going on. And I heard
12 it here again today. We want teachers to do
13 successful things as a prescription from other
14 teachers who are doing it well. I have close to
15 50 years of seeing effective teachers. And I'm
16 telling you that no two of them were effective in
17 the same way. And so anything that imposes
18 teacher uniformity makes me itch.

19 I think that we can look at
20 unanticipated consequences very nicely by looking
21 at what's happened since No Child Left Behind was
22 passed. The assessment system that we have is

1 supposed to inform students and their parents of
2 where they are academically and it is supposed to
3 identify which schools are successful and which
4 schools are not. It doesn't do a particularly
5 good job of either in my opinion. But it's
6 unanticipated consequence is that it did publicize
7 to parents exactly where their student was on the
8 normal curve. It, therefore, became practice in
9 districts, and in particular mine is what pushed
10 me into retirement ultimately I think, were people
11 were being trained to identify the normal curve
12 band equivalent of students that they could
13 improve most quickly. In other words, people
14 talked about the biggest AYP bang for the buck.
15 And that drained the resources into the group that
16 we know in California as in the basic category
17 because they had a lesser distance to go to
18 proficient than anyone else.

19 Money on groups that were below basic or
20 far below basic was regarded as essentially wasted
21 because it was a much bigger instructional lift to
22 get those people anywhere close to proficient.

1 This has allowed all of the resources or the
2 majority of them to be drained to this group where
3 the improvement was most easily achieved. It has
4 been drained in the form of the purchase of
5 glossy, success-guaranteed, scripted programs
6 which leads, in my opinion, to a decline in
7 teacher abilities, in curriculum knowledge,
8 planning for instruction, and practice of a
9 variety of instructional strategies. Give me a
10 script and I am telling you I would have to either
11 subvert it or make it mine or slit my wrists
12 within a month. And it is not anywhere near
13 useful to children who all learn in different
14 ways.

15 The horrible effect on special ed is the
16 resources have been largely drained from
17 mild/moderate -- children classified as
18 mild/moderate, but also -- and we should be
19 terribly worried about this -- it has been drained
20 from the gifted who also need and deserve help.
21 Many of them I might point out to you have
22 specific learning disabilities which need to be

1 addressed. And I just resent terribly this. And
2 so I am pleading with you that whatever you set
3 up, you do it so that teaching is improved because
4 that is the only way that learning will be
5 improved.

6 MS. WEISS: Thank you. Next?

7 MS. VILLEGAS: Hi. My name is Rosa
8 Villegas. I thank you for giving me the
9 opportunity to be here. I'm representing LAUSD,
10 Los Angeles, California. I'm also on
11 Multicultural Advisory Committee, a chair of
12 LAUSD, a Special Education Advisory Committee, and
13 I'm also representing North Lake County Regional
14 Center and their board.

15 The reason that I'm here, I'm extremely
16 concerned about the Race to the Top programs
17 because I see it as a parent. My concern with the
18 Race to the Top program is that no matter what
19 people say, this here -- our students with
20 disabilities are extremely discriminated because
21 the services are not being provided. And services
22 that are stated on the IEP, they do not get

1 addressed. And with this assessment program I see
2 a lot of disadvantage to our children.

3 And also there is another concern that I
4 have that I want to say about charter schools.
5 Charter schools will never take our children,
6 especially my child that has significant delays
7 with disabilities. Charter schools will only take
8 children that are highly gifted. Because charter
9 schools do not want the grades to be lowered. So
10 I'm extremely concerned about that. So I would
11 please -- I would like for you to take that into
12 consideration. I'm extremely concerned. And I'm
13 speaking here on behalf of all the parents that
14 have children with disabilities because we love
15 our children and our children are the most of
16 everything. I'm asking you to please give the
17 opportunity to our children. They have a future.
18 But without giving them the opportunity, they have
19 no future. We just have to believe and trust in
20 them and provide the services and the quality
21 education that they deserve.

22 But also, the teachers, they need to be

1 trained. Because if they don't have the
2 experience how are they going to teach our
3 children? So, please, whatever you can do to help
4 the teachers and parents with the special
5 education I would truly appreciate it.

6 Thank you for having me. Thank you
7 again.

8 MS. WEISS: Thank you.

9 MS. GREEN: Good afternoon. My name is
10 Nancy Green and I'm the executive director of the
11 National Association for Gifted Children, the
12 nation's largest organization dedicated to
13 supporting gifted learners.

14 Our nation as a whole has long neglected
15 the needs of our highest potential students.
16 Several issues are to blame, including the lack of
17 a comprehensive national gifted education policy,
18 a piecemeal system of identification, and services
19 that vary sharply between and within states and a
20 national focus too heavily weighted toward
21 achieving proficiency alone.

22 For many reasons, the status quo is not

1 serving our nation and our more than three million
2 academically gifted children. On behalf of these
3 students, I offer the following recommendations
4 regarding the Race to the Top Assessment Fund.
5 NAGC vigorously supports the intent of the
6 assessment program. To improve it though, we
7 recommend a special focus on the precision of
8 assessments for students at the upper end of the
9 achievement spectrum and the use of
10 computer-adaptive assessments for this population
11 of students. Empirical and anecdotal evidence
12 demonstrate that the assessment and accountability
13 provisions in NCLB have had negative consequences
14 for our nation's advanced learners. Indeed, the
15 assessment and accountability system created a
16 much-needed focus for educators to address the
17 needs of students who do not read or perform
18 mathematics at grade level.

19 Unfortunately, this incentive structure
20 has driven educational systems to neglect students
21 performing above the proficient level in reading
22 and math. As a result, the academic progress of

1 top students has languished during this same time
2 period. The current system has also helped widen
3 dramatically the racial achievement gap between
4 students on the higher end of the achievement
5 spectrum. As districts focus on their struggling
6 learners, new research shows that the achievement
7 gaps in reading and math skills between
8 kindergarten and 5th grade grow roughly twice as
9 fast for students who begin school with scores one
10 standard deviation above the mean as for those who
11 begin one standard deviation below the mean.

12 This trend is especially stark between
13 high achieving minority and white students. Also,
14 half of lower income students classified as high
15 achievers in math early in their schooling lost
16 this status later in elementary school. Clearly
17 we need to do better. Focusing grants on the
18 design and quality of assessment systems to
19 support improved teaching and learning, rather
20 than just on accountability policies, is a
21 critical improvement. However, to achieve the
22 Race to the Top program goals, funded projects

1 must encourage the development of assessments that
2 accurately measure learning gains for students
3 across the entire achievement spectrum.

4 Most of the discussion among
5 policymakers focuses on the reliability of
6 assessments for students with disabilities, which
7 has sparked innovation in the development of new
8 assessments. By contrast, the same innovation has
9 not developed for gifted learners who also require
10 assessments that are able to accurately measure
11 their achievement, which frequently exceeds grade
12 level standards. NAGC strongly urges the
13 Department to include a priority for valid and
14 reliable measures of student achievement and
15 growth at the upper end of the performance scale.

16 We understand that the new Common Core
17 Standards are organized in bands that cover
18 multiple grades. We encourage the corresponding
19 grade level assessments to also cover multiple
20 grades. This strategy may help mitigate the
21 ceiling effect which so many gifted students
22 experience when taking standardized tests.

1 Computer-adaptive assessments, which are growing
2 in popularity, are another issue of special
3 interest. The promise of these assessments is to
4 be able to pinpoint students' ability levels with
5 test questions based on prior test performance.
6 With this data, educators will have the important
7 information they need to differentiate and adapt
8 instruction to meet student needs. By determining
9 students' mastery of a subject, teachers can
10 implement pedagogical strategies, such as academic
11 acceleration and flexible ability group to ensure
12 that gifted students meet not only a grade-level
13 standard, but also their individual learning
14 potential.

15 NAGC is concerned that an emphasis on
16 measuring student performance against only grade
17 level standards will continue to harm some gifted
18 students by preventing them from reaching their
19 full academic potential. Educators will continue
20 to be driven to teach to the middle, rather than
21 maximizing student learning gains. Together we
22 must champion a new movement that strives for

1 excellence in education rather than the mediocrity
2 that results from merely asking students to reach
3 proficiency. We must challenge educators and the
4 public to abolish preconceived notions of what a
5 child at a certain age ought to know and instead
6 provide students with the academic support to
7 exceed their own and our expectations.

8 Thank you.

9 MS. WEISS: Thank you.

10 MR. RING: Good afternoon. My name is
11 Bill Ring. I'm a parent and parent advocate. I'm
12 a director of TransParent and here today on behalf
13 of the Parent Collaborative as its first vice
14 chair of the Los Angeles Unified School District's
15 parent collaborative.

16 I'd like to spend what time I have today
17 addressing the last question posed by the
18 Department. So are there other issues that need
19 additional focused research? Initiatives come and
20 go. Who thinks about asking students how they
21 like to learn? Discussions about standards and
22 assessments reminded me of something I contributed

1 to. An online dialogue about the California
2 Master Plan in 2002. At the time I had just
3 helped an 8th grade student in his research for a
4 paper he was doing on education. His last
5 question to me was this, "How do people decide
6 what kids need to learn?"

7 I added two questions to that. Who are
8 those people? And why do they have such
9 authority? So who exactly is setting the
10 standards? Regardless of the answer, I don't
11 think that raising standards in the existing
12 system is going to change it into anything other
13 than what it is. I said that and other parts of
14 that, which I don't have time to get into eight
15 years ago. Here we are today. And for me, the
16 same questions apply. The same concern over
17 narrowing the curriculum. The same concern
18 expressed by others. And the question -- are we
19 measuring the wrong end of the student?

20 So you ask, are there other issues that
21 need additional focused research? Some parents
22 with students at our high schools are pleased

1 because teachers like their children, yet some of
2 these same students are receiving 3.5 GPAs and
3 higher and still can't pass the state's high
4 school exit exam. This should be exhibit one in
5 the argument that parents don't know what
6 questions to ask often or what to look for. And
7 worse, that the school and the district and state
8 have failed to build the requisite capacity in the
9 school community to fully comprehend standards and
10 assessments and what the students need to and are
11 expected to learn. It also exposes grade
12 inflation and begs a broader set of assessment
13 questions for me.

14 Parents are not generally welcomed as
15 partners in policymaking and decision-making in
16 L.A. Unified. And elsewhere I would wager.
17 Perhaps as a provocative example, when it comes
18 time to talk about wages and benefits and working
19 conditions for teachers who are teaching our
20 children and administrators who are running our
21 schools, union representatives sit down with the
22 district and talk about such things, but the

1 parents are not included in the dialogue. We and
2 our children are the clients. We need to be
3 included in these discussions which are directly
4 connected to student achievement and budget.

5 All the talk about teacher quality,
6 employee evaluation, seniority and the like raises
7 some questions for parents about how districts
8 deal with collective bargaining and what is truly
9 in the best interests of students and children.
10 So schools that -- schools don't always educate
11 parents to become equal partners in shared
12 decision-making and often they don't want us
13 involved at all.

14 Last May, I was part of a group of
15 parents presenting to the California State Board
16 of Education about the inadequate training of
17 parents, which has led to a lack of capacity and
18 frankly, a lack of trust. Parents routinely see
19 little monitoring and evaluation of decisions to
20 track results. And as we all know, you cannot
21 manage what you don't measure. Schools often fear
22 having candid conversations around data, and when

1 they do weigh it in there is sometimes, and often,
2 difficulty in having a common conversation with
3 all members of the school community.

4 Practice is a reflection of what we
5 value. I am here to make the case for the need to
6 make the case for the need to include parents and
7 secondary students in the assessment and
8 evaluation of teachers, principals, and other
9 administrators -- and yes, even superintendents.
10 We, parents, are consumers like everyone. But in
11 this case, we're consumers of educational services
12 provided to our children. In the commercial
13 marketplace when products and services are
14 introduced, it is typical that consumers are
15 engaged in such development. Their opinions are
16 sought, surveys are taken, user groups are
17 assembled and polled, and certainly in the
18 standards and assessment business providers must
19 recognize that users and developers work together
20 to seek solutions.

21 Where in Race to the Top is the
22 commitment and specific acknowledgement and

1 recognition of the need to build capacity in our
2 parents and in our broader school communities? I
3 would like to see pilot of a school inspection
4 system. An inspector, if you will, as suggested
5 by Californian Stephen Blake. Blake speaks of a
6 state-specific inspection framework. But perhaps
7 this idea could be introduced as a part of or in
8 conjunction with Randy Bennett's school-based
9 model for assessment innovation we heard about
10 today, even while recognizing Tom Vander Ark's
11 conclusion in 2005 that giving failing schools
12 autonomy is a bad idea. We must change the
13 culture to embrace, train, and trust parents and
14 families as partners in this work.

15 Here are a couple of additional
16 candidates for research by the Department and
17 consortia. The draft of California's Race to the
18 Top plan dated January 5, 2010, included a bullet
19 point to increase choice and empowerment of
20 parents and stated that California is encouraging
21 parent- school partnerships to include student
22 achievement. But it goes on to say that the

1 parents in the lowest achieving schools as defined
2 by Race to the Top will have the option to apply
3 to transfer to a higher performing school in
4 another district. Is opting out an example of
5 partnership? Instead of abandoning the system we
6 should be driving it.

7 Scott Marion asks here this morning do
8 other forms of school reform work better? Well,
9 parents with students in failing charter schools
10 have already pulled the trigger, yet the schools
11 are still failing their children. Now what?

12 MS. WEISS: Thank you, sir. Your time
13 is up.

14 MR. RING: Thank you.

15 MS. WEISS: Thanks very much.

16 MS. DARLING-HAMMOND: Hi. I'm Linda
17 Darling- Hammond. I'm a professor at Stanford
18 University where I've conducted research on
19 assessment systems in the United States and
20 abroad. And I've been involved in the development
21 of student and teacher assessment programs in a
22 number of states. I'm currently leading a project

1 which synthesizes what we've learned in the U.S.
2 and other nations about how to design, score, and
3 manage performance assessment systems that are
4 valid, reliable, feasible, and affordable. There
5 are a set of papers from that project that will be
6 available by March, including one by Jamal Abedi
7 on Performance Assessments for English Language
8 Learners.

9 And along with my comments today, I will
10 be submitting several of these papers from that
11 set of studies which describe the operations of
12 performance assessment systems in high achieving
13 countries; technical advances in performance
14 assessments in the United States over the last 20
15 years that allow us to create more reliable,
16 valid, and affordable assessment tools and cost
17 estimates that illustrate how to make rich
18 performance-based assessments affordable in an
19 integrated system that supports teaching and
20 learning.

21 And in this very brief time I just want
22 to make three points. First, creating a system

1 that routinely incorporates performance-based
2 components in both the formative and summative
3 aspects of our assessment system is essential to
4 improving the quality of teaching and learning in
5 our schools. I'm remaking the validity argument
6 that Scott Marion started us with this afternoon
7 and Laurie Weiss, I think, ended us with, that
8 whatever we do, it's critical that we actually ask
9 students to show that they can write and
10 communicate, research and investigate, be problem
11 framers and solvers, thinkers and producers of
12 ideas and products. And if we don't do that I
13 actually think we'll see very little gain from the
14 common standards or new assessments.

15 Even in our better schools. I think our
16 students in our better schools are actually doing
17 much less intellectually rigorous work than
18 students in high achieving nations abroad where
19 these kinds of activities are substantial and a
20 growing part of the assessment systems. And so I
21 think we need to take a close look. Many of the
22 questions that were raised this morning are

1 answered in very innovative and creative ways in
2 places like Finland, the Netherlands, Hong Kong,
3 Singapore, Australia, New Zealand, and so on. And
4 the paper that I'm submitting outlines the
5 operations of systems that reflect many of these
6 questions.

7 I was the person who coined the term
8 "through- course" assessment in Boston on the
9 assessment panel there, to describe what many of
10 these nations do and to debate the point that
11 end-of-course exams as we call them here, are not
12 end-of-course exams in other countries. They are
13 through- courses assessments that have multiple
14 components.

15 And I'd like to clarify what I meant by
16 that. And it was really in line with what I think
17 Jeff Nellhaus and Lizanne and others pointed out
18 that they'd like to see. That the assessments are
19 curriculum-embedded tasks, like science labs and
20 investigations, written papers in various genres,
21 research, applied mathematics tasks, computer
22 programs that students develop. They're not mini

1 tests that are used to predict the next test.

2 So the goal is to be sure that students
3 have rich learning opportunities, teachers have
4 rich evidence-basis and full ranges of standards
5 are measured. The tasks really represent central
6 modes of inquiry in the disciplines. They're not
7 overly specified in that, for example, if you look
8 at the controlled assessments in the GCSE or the
9 tests and tasks in England or the tasks in the
10 International Baccalaureate system, they may say
11 compare two or three novels you've read around a
12 particular theme, but they don't predict the
13 novels you've read. They allow for flexibility
14 while having common standards. So this I think is
15 worth taking a hard look at particularly around
16 the high school assessment systems, which I don't
17 have time to talk further about now.

18 Second, creating such a system is much
19 more feasible than it once was. There have been
20 substantial advances since the 1990s in our
21 understanding of how to design and score
22 generalizable tests that can be scored in

1 comparable ways. And some of these advances are
2 described in one of the papers that I'll submit.
3 We can enact more consistent, effective, and
4 affordable scoring.

5 And finally, such a system is both more
6 cost-effective and can be made more affordable
7 through economies of scale offered by consortia,
8 new uses of technologies, and thoughtful
9 approaches to teacher scoring. The cost
10 implications of various approaches will be
11 described by John Olson, who will be speaking in a
12 few minutes in a paper that we included in that
13 set and will be submitted to you.

14 Finally, among the several consortia
15 that are currently forming there are three that
16 are already beginning to work together to
17 conceptualize and implement a balanced assessment
18 system with performance-based components that does
19 take advantage of both formative elements and the
20 affordance of technology. So I think there's
21 great energy in the field which you have really
22 stimulated by this wonderful process over the last

1 few months to move forward in significant ways.

2 MS. WEISS: Thank you.

3 MR. PATZ: Good afternoon. And thank
4 you for assembling such an informative day and for
5 inviting participation.

6 My name is Richard Patz. I'm the vice
7 president for research and product development at
8 CTB/McGraw-Hill. CTB was founded in 1926 and has
9 been building nationally standardized assessments
10 for decades. And we have done a great deal of
11 work also with states currently under the existing
12 No Child Left Behind program. I mention that to
13 say that there are -- if you look across the
14 states and across the different vendors doing
15 different things in states -- there are some very
16 useful, promising, innovative practices going on.
17 There's the use of writing assessment now with
18 artificial intelligence scoring; computer-based
19 tests being used in increasingly useful ways;
20 interesting technical work in scaling and growth
21 models, just to name a few.

22 So when we think about the future it's

1 clear that there are features of the current
2 accountability system that we would like to leave
3 behind, but there are also some things that we
4 should examine so we can see what we can leverage
5 that's been successful and to not assume that we
6 have to build everything from scratch. In any
7 case, we care deeply about the evolving
8 conversation about education and testing reform,
9 and I appreciate the opportunity to contribute.

10 I have four quick observations in light
11 of the questions that were raised in advance of
12 the meeting, specifically on the general and
13 technical assessment matters. First, summative
14 assessments -- those that have real consequences
15 for students, schools, teachers -- they must be
16 standardized and validated. They should be built
17 in compliance with the standards for educational
18 and psychological testing published at AERA, APA,
19 and NCME; reviewed and approved by appropriately
20 credentialed technical advisory committees;
21 administered with appropriately high levels of
22 test security.

1 Now, assessments meeting these
2 requirements may measure common course standards;
3 they may include diverse item types soliciting
4 complex student performance. And that can be done
5 in technically credible ways. The Yen and Ferrara
6 paper from '97, for example, looked at the
7 Maryland State Performance Assessment Program,
8 which was very innovative in some ways and very
9 technically well done. And these programs can be
10 designed in innovative ways as well. And I
11 published a paper on that that I will be happy to
12 share -- in 2006 measurement.

13 Secondly, if the through-course
14 assessments under consideration by the Department
15 could be built and used as interim assessments and
16 not used for the summative purposes that require
17 standardization, then important goals might be
18 achieved sooner. Such innovative interim
19 assessments can involve complex tasks and
20 innovative item types. They can be thoroughly
21 researched, scaled if appropriate depending on the
22 purpose, linked to summative tests through

1 research. But free of the burden of
2 standardization and validation that's required for
3 the summative uses, the test could be used much
4 more flexibly and in a variety of ways, including
5 being administered at different times and other
6 things that we want to do. Test security would be
7 much simpler and teacher professional development
8 could be enhanced and you would not lose the
9 pedagogical value of a comprehensive summative
10 assessment at the end of the course.

11 Thirdly, I think it's important -- it is
12 more important that tools for formative and
13 interim assessment integrate seamlessly with
14 instructional practice than it is for those tools
15 to integrate seamlessly with the summative tests.
16 The tools that enable teachers to utilize student
17 performance data in real-time or near real-time
18 and adapt instruction accordingly are distinctly
19 different from the tools required to securely
20 administer a high-stakes, large- scale assessment.
21 So, although it's possible to create one system
22 incorporating all these tools, separate, but

1 connected systems may be more feasible and more
2 optimal.

3 Finally, opportunities to improve K-12
4 assessment abound. And in this endeavor you'll
5 find some of the best talent available to help
6 guide this improvement resulting in testing
7 organizations as you saw today. Finding ways to
8 engage this talent, to learn from past successes
9 and failures, and to leverage existing
10 capabilities will be critical to designing future
11 assessment programs that are truly innovative,
12 technical defensible, and feasible.

13 So CTB/McGraw-Hill looks forward. We
14 support the goals of the Common Core Standards
15 Initiative and the Race to the Top Program, and we
16 look forward to opportunities to support the U.S.
17 Department of Education, and the states and their
18 consortia, and the organizations that are
19 advancing these important goals.

20 Thank you.

21 MS. WEISS: Thank you.

22 MR. FOSTER: Good afternoon. Please

1 excuse my back.

2 My name is Dr. John Foster. I am the
3 president and CEO of NOCTI. That's formerly the
4 National Occupational Competency Testing
5 Institute. NOCTI is a not-for-profit assessment
6 firm that collaboratively serves the secondary
7 community college and university levels of career
8 and technical education. We very much appreciate
9 the focus on the word technical. I have a
10 different spin on that than most of the other
11 speakers. We use online, paper and pencil,
12 performance, interactive items, and text-to-speech
13 in our work. And we really appreciate the focus
14 that you brought on the technical pieces of
15 assessment under the RTT discussions.

16 I'm going to ask you to stretch your
17 perspectives a little bit and think about lessons
18 that can be learned in kind of a parallel universe
19 -- that of career and technical education. I'm
20 going to talk about three points, basically,
21 technical assessments -- or three differences,
22 really, in terms of technical assessments as we

1 know them. And I talk technical in terms of
2 technical related to the occupational world.

3 Technical assessments are of an
4 immediate and critical nature due to their impact
5 on assuring technical competence of the incumbent,
6 as well as the incoming workforce. A short
7 competence is desperately needed as employers
8 scramble for talent coming out of the current
9 recession and to maintain this country's global
10 competitiveness.

11 Second, technical assessments and
12 certifications are dynamic. They need to reflect
13 continual upgrades and technical skills,
14 processes, materials, and equipment, yet the same
15 assessments must meet the psychometric standards
16 of their academic counterparts with a more static
17 content.

18 Thirdly, working on technical
19 assessments, the work with technical skill subject
20 matter experts presents numerous challenges.
21 These subject matter experts typically can't
22 commit to time to design a test; they sometimes

1 represent regional differences; and those kinds of
2 things must be compensated when you're talking
3 about a nationally standardized -- be it voluntary
4 or otherwise -- sort of an assessment.

5 So innovative test development
6 approaches must be implemented. All these issues
7 add up to additional funding and additional
8 expense, and the Department of Labor's Funding
9 Initiative that embraces a competency modeling
10 format -- think performance modeling -- and the
11 recent announced grant to establish a certifying
12 exam for those entering new fields like the
13 Medical Information Systems are exceptional steps.
14 However, they are not enough. And what's more,
15 these initiatives are not very familiar to the
16 General Education Committee -- Community, sorry.

17 We have seen many positive assessment
18 decisions being made across the nation and have
19 begun to see cross state collaboration models. We
20 have also seen some statewide technical
21 articulation of secondary or postsecondary
22 programs -- both ways. Secondary to community

1 college; community college to university. And
2 these are based on assessment vehicles or a
3 combination of assessment vehicles. We believe
4 that some states are already picking up the Race
5 to the Top challenge.

6 We also have witnessed our share of
7 concerns. Most of these concerns have to do with
8 adequate and equitable resources. As a brief
9 illustration, we are contracted by one state to
10 develop dynamic technical skill attainment tests
11 in 50 occupational areas. Our contract was
12 approximately \$1.2 million over a 3-year period.
13 During that same period, another third party
14 developer received \$23 million to develop three
15 static academic assessments. Quick division shows
16 resources roughly 108 times higher on development
17 of each academic test compared with each technical
18 test. When policymakers indicate the assessment
19 of technical competence is supposed to ensure the
20 maintenance of this country's infrastructure and
21 sustainability of our economic prosperity, it is
22 difficult to reconcile inequities of this

1 magnitude.

2 NOCTI has demonstrated a willingness to
3 help with the ongoing assessment efforts. We have
4 demonstrated that willingness by participating as
5 a partner in the National Research Center for
6 Career and Technical Education, by working with
7 the Office of Vocational Adult Education here at
8 the federal level, and also participating in data
9 quality institutes. We have also corresponded
10 with the Secretary's Office, basically
11 underscoring that we would be happy to contribute
12 our national perspective and expertise.

13 Like our experts earlier today, our work
14 requires in-depth knowledge of the process of
15 building standards and following those with a
16 test. We understand the need for item analysis,
17 documented varies of validity and reliability, and
18 we know what it takes to develop a nationwide test
19 -- whether it's online or otherwise -- with proper
20 weightings and ratings. We hope to be of service
21 to the educational community, particularly the CTE
22 community. And we would encourage equitable

1 funding regarding the CTE assessment initiatives.
2 And certainly encourage innovative approaches as
3 we all move forward.

4 In closing, I just want to make three
5 points. One, remember CTE. According to the
6 NELS, over 90 percent enrolled at the secondary
7 level participate in CT; that's a big number. CTE
8 teachers represent the second largest group of
9 teachers in all high school subjects -- all
10 secondary subjects. Second, take advantage of our
11 and other members' of the National Research Center
12 on Career and Technical Education's technical
13 expertise.

14 MS. WEISS: Thank you.

15 MR. FOSTER: And early, encourage
16 innovation. Thank you very much.

17 MS. WEISS: Thank you so much.

18 MR. OLSON: Hi. I'm John Olson,
19 representing the Assessment Solutions Group. My
20 topic is "The cost of new higher quality
21 assessments: A comprehensive analysis of the
22 potential costs for future state assessments."

1 In our opinion, the Race to the Top and
2 Common Core Standards are two key initiatives for
3 improving education. And states will have an
4 opportunity to receive some much needed resources
5 and assistance to help them make important
6 improvements to their assessments and create a
7 higher quality assessment system. State
8 assessments need to be improved so they do a
9 better job of measuring the critical skills
10 students will need in the 21st century, are
11 integrated into the curriculum, help students
12 learn, and provide teachers opportunities to
13 develop new instructional strategies. Higher
14 quality summative and interim assessments need to
15 be well designed, include more constructive
16 response and performance items, and be affordable
17 to states. It is critical that states be able to
18 afford the long-term ongoing administration costs
19 associated with any new assessment system. Should
20 states design a new assessment system without a
21 thorough review of costs and affordability prior
22 to project implementation, the sustainability of

1 any improvements to the current system of
2 assessments could be in jeopardy.

3 The Assessment Solutions Group is
4 conducting a study of the costs for various
5 assessment designs, including different models and
6 options for higher quality state assessments. We
7 are collaborating with Dr. Linda Darling- Hammond
8 and her colleagues at Stanford University and are
9 in the process of completing a research paper
10 which focuses on the design and costs of higher
11 quality, balanced assessment systems that will
12 meet the federal characteristics of the next
13 generation assessment at affordable price to
14 states. The research we are conducting for this
15 study and the preliminary results reported in our
16 draft paper demonstrate that under the right
17 conditions states can dramatically improve their
18 assessments and develop a higher quality
19 assessment system at an affordable cost. However,
20 states must be careful to design an efficient
21 assessment system and understand its ongoing
22 administration costs, as well as their state

1 budget allocations prior to committing to an
2 innovative, high quality assessment and
3 implementing it in their state. Professional help
4 in these areas is highly recommended.

5 Based on our detailed cost analysis, the
6 overall finding is that a higher quality
7 assessment, one that meets the criteria spelled
8 out in the USED design of assessment systems for
9 Race to the Top can be done for about the same
10 money as is currently being spent for traditional
11 state assessments today. Our draft paper provides
12 more details of the findings from our study, some
13 of which follow.

14 Our preliminary results indicate that
15 such a higher quality assessment system can be
16 developed that has increased numbers of
17 constructed response items, performance-based
18 tasks, teacher scoring of these items, as well as
19 other features and will cost roughly the same, or
20 in some cases less than what a typical state pays
21 for an assessment today. It was found that the
22 development costs for a new, high quality

1 assessment is relatively inexpensive relative to
2 the total cost of the assessment. The ongoing
3 administration costs are by far the largest
4 component of any assessment system. A key factor
5 in the sustainability of new and improved
6 assessments and whether or not states can adopt
7 and use them will be the ongoing administration
8 costs that need to be managed. It also is
9 critical that the next generation of assessments
10 be designed so they're affordable on an ongoing
11 basis to states.

12 Of the cost reduction strategies that
13 were examined, teacher scoring with and without a
14 stipend and information of consortia were the two
15 most important means of controlling overall cost.
16 In order to share costs across states, it will be
17 important to have states participate in assessment
18 consortia to share the overhead associated with
19 development and management of assessments. To
20 help control costs, teachers will need to score
21 the performance items. This practice is used in
22 many high-performing countries and helps build

1 teacher expertise and understanding of student
2 performance expectations. The use of
3 computer-based technology is recommended, but not
4 required to achieve the savings needed to
5 implement a higher quality assessment at current
6 prices. Online assessments can save significant
7 costs versus paper and pencil tests.
8 Additionally, distribute scoring of responsive
9 items can save costs. And finally, the use of
10 computerized artificial intelligence scoring holds
11 significant promise in making high-quality
12 assessments more affordable. A clear and
13 efficient assessment design must be created and
14 specified in an equally clear RFP. Such a design
15 can minimize costs, particularly development and
16 printing logistics costs.

17 In summary, we believe that states need
18 to work together to create the next generation of
19 high quality assessments. More details are in the
20 paper we are providing to the Department with
21 detailed cost estimates, too.

22 Thank you.

1 MS. WEISS: Thank you.

2 MS. KAPINUS: Good afternoon. My name
3 is Barb Kapinus. I am a senior policy analyst in
4 the Education Policy and Practice Department at
5 the National Education Association.

6 And first I want to thank you for
7 bringing together some incredible experts with
8 very impressive ideas about assessment. I am much
9 -- I'm very optimistic about the possibilities of
10 this funding after listening to the people that I
11 know you're listening to.

12 Because I am speaking on behalf of the
13 National Education Association and our teachers,
14 I'm going to be emphasizing the importance of
15 considering the role of teachers in the
16 development, implementation, and improvement of
17 assessment systems, rather than sort of looking at
18 the whole system. And rather than looking at some
19 of the technical aspects and things that have been
20 covered very well by other people who know much
21 more about those things than I do.

22 Because formative assessment is a

1 teacher tool for classroom assessment and
2 planning, I want to particularly support its
3 importance as part of a complete assessment system
4 that leverages education improvement. NEA wants
5 every child who attends a public school to receive
6 a complete and worthwhile education. The key to
7 creating and sustaining such schools is the
8 development of accomplished teachers who
9 continually strive to improve their professional
10 practice. Simply administering assessment after
11 assessment by itself will not accomplish the goals
12 of providing schools -- of developing schools who
13 provide these kinds of educations.

14 This has been stated by several of the
15 people who have testified here in the last few --
16 across all of your hearings. However, I do
17 believe that effective assessment systems, when
18 they're accompanied by other resources for the
19 well-being of students, can be a tremendous factor
20 in the improvement of quality education. In order
21 to construct such assessment systems, it's
22 necessary to examine the goals of the education

1 assessment and put aside much of what we've been
2 doing in education assessment in the past. And I
3 have heard several people and read their testimony
4 to that effect.

5 Current assessment practices are linked
6 to the goal of attaining specific scores on highly
7 constrained tests in order to prod states,
8 districts, and schools into doing a better job.
9 For accountability related to the NCLB, many
10 states use standardized tests that are relatively
11 easy to administer and score, but have little
12 connection to what a student learns as part of a
13 quality education or a quality instructional
14 program. They're really not designed to ensure
15 that kids have learned. The process of assessment
16 student learning should be directly linked to
17 improving student learning. Current large-scale
18 assessments are only indirectly linked to that
19 goal.

20 Much money has been poured into those
21 assessments, but there's no research evidence that
22 they actually improve student learning. Improved

1 scores on those tests sometimes reflect other
2 things such as drills and focusing on very
3 specific things that are on those tests, rather
4 than a really deep understanding of the concepts
5 that have been taught and the ability to apply
6 complex skills.

7 To promote learning in the classroom we
8 need better assessment tools that directly impact
9 student learning -- what are known as formative
10 assessments. And by formative assessments we mean
11 assessments that provide detailed information
12 about student learning directly and clearly to
13 both students and teachers. They allow for
14 re-teaching and re-doing of tasks so that the
15 point of these things is not to get a grade or a
16 score, but the point is to make sure that the kids
17 get it; that they can do what we want them to do;
18 that they know what we want them to know.

19 These assessments also provide
20 opportunities for teachers to develop essential
21 skills and deep knowledge of both standards and
22 their students. There is an urgent need to

1 provide resources for professional development
2 related especially to these formative assessments
3 that according to Randy Bennett and several others
4 are part of assessment of, for, and as learning --
5 an essential, yet neglected component of
6 assessment systems.

7 The NEA and its teachers stand ready,
8 willing, and able to give input and to support
9 efforts to develop the quality assessment systems
10 that you're seeking to fund.

11 MS. WEISS: Thank you so much.

12 MR. RICHARDSON: Good afternoon, ladies
13 and gentlemen. My name is Walter "Waddles"
14 Richardson. I'm a parent, advocate, and I'm from
15 the Los Angeles Unified School District.

16 I want to thank the committees that
17 spoke this morning and this afternoon regarding
18 the assessments. I feel that they spent an
19 in-depth amount of time in gathering all the
20 information. However, I do wonder if they went
21 for enough in-depth as far as with their
22 information. I kept looking to hear them

1 represent all of the students of ethnicity and I
2 never heard that. And I'm wondering if we're
3 going to do Race to the Top to be inclusive of all
4 students, then we have to have data on all of the
5 students that is in your particular school or your
6 state or your school district. And so that kind
7 of bothers me and puzzles me.

8 Another thing that puzzled me was the
9 fact that I had looked around and I was hoping
10 that I would have seen at least one
11 African-American sitting up on these committees
12 today, and I didn't see any. And so it kind of
13 bothered me because I was wondering what type of
14 message was it sending. Or it was just an
15 oversight? Are you saying there isn't a qualified
16 African-American that is able to sit on the
17 committees to give input? I wondered also at the
18 same time saying, well, what would America be like
19 without African- American in it? What would the
20 school district system be like without
21 African-American in it? I think we've made a lot
22 of contribution.

1 The only thing that I'm saying is that
2 if we're going to come up with a Race to the Top
3 assessment, we need to be inclusive of all of the
4 ethnicities to ensure that we are addressing the
5 specific needs of all of the students. I think
6 and hope that is the idea of the whole process.

7 Also, I'd like to reintroduce myself.
8 Again, I'm Walter "Waddles" Richardson. I'm the
9 second vice chair of the District Advisory
10 Committee -- DAC -- in the Los Angeles Unified
11 School District. I represent some 461,128
12 identified students. Those schools -- and we have
13 590 identified (inaudible) schools. In those
14 schools, these students are low performing. Not
15 all of them, but they have some type of
16 disability. And so I'm very much concerned that
17 we reach and reach out and grab -- get those
18 disability -- we have special ed students that are
19 mainstreamed back into the regular program. We
20 have left students. L.A. Unified has, what, 200
21 and some identified left students. And many of
22 them, about 800 (inaudible) their language is

1 speaking in Spanish. And so we got to make sure
2 that we be inclusive of them.

3 And then we look down and L.A. Unified
4 -- the State Board of Education went and did a
5 survey of 500,000 identified -- 500,000
6 African-American males and found out they
7 represented 16 percent of the school population in
8 the state of California. And out of that 16
9 percent, 11 percent of them had been put off into
10 special education. That's sadful. It's sinful.
11 All they did was identify some kids that had some
12 type of erratical un-behavior and problem. And
13 then they said you know what? We're putting you
14 in special ed. That's sinful. That cannot be
15 accepted, ladies and gentlemen. We have to do a
16 better job than this.

17 Suppose we had took Eddie Murphy, Mr.
18 Jordan, Michael Jackson, and some of the many
19 others. One dancing, one shooting the hoops and
20 doing all kinds of other things. And look at all
21 of them. They all turned out to be very
22 successful. So we cannot give up on our children.

1 What I'm saying to is we have to look -- if we're
2 going to be dealing with the consortium -- I
3 helped to write consortiums in L.A. Back in --
4 there were nine consortia in L.A. -- in L.A.
5 County, and I helped to write a plan that the
6 district never did really use or implement.
7 However, it's coming back into place now.

8 What I'm saying is that when we come up
9 with these data and things, we need to be sure
10 that we're using the data that we are going to
11 actually attract and be able to address the whole
12 child. I'm finding many places I go and talk with
13 people all over the land -- we're speaking about
14 -- well, we're not doing that part this year on
15 the children. Well, you can't address a whole
16 child because I got a whole child that's in the
17 10th grade. She's an honor roll student because
18 dad is in the classroom two or three times a week.
19 I'm in the principal's face. I'm in the staff
20 development. We had some teachers where they
21 called the police on the parents because they sit
22 in staff development. Okay?

1 Thank you, ladies and gentlemen.

2 MS. WEISS: Thank you very much.

3 MS. GENDRON: Good afternoon. I'm Susan
4 Gendron. I'm commissioner of education for the
5 State of Maine. I'm also president of the Council
6 of Chief State School Officers, and I'm also proud
7 to be one of the NECAP states. And we've had
8 great opportunities as a new member to that
9 consortium.

10 But I'm here today to also thank all the
11 experts and all of you for the great questions
12 you're asking us. The states really do stand
13 ready to embrace this challenge and opportunity
14 for us to create a truly balanced assessment
15 system for all students in our country.

16 As we've all been working and preparing
17 for Race to the Top, many of us have begun the
18 conversation about what are the right consortiums?
19 Who are the right partners? You're going to see,
20 or the peer reviewers will see as they look at
21 those applications, there are six consortiums that
22 have begun to emerge. Three of those consortiums

1 are having very detailed conversations. And I'm
2 one of the chiefs who has been advocating for a
3 balanced assessment approach. And as of today,
4 there are 38 states who have signed onto that
5 consortium approach.

6 And what I'd like to do is to share with
7 you what our thinking is to date. And I've
8 learned a great deal today that will help to
9 inform our conversations. We really do want to
10 see a thoughtful, integrated system around
11 balanced assessment that will focus -- you talked
12 a lot today about standards, curriculum,
13 assessment, instruction. Teacher development has
14 got to be a critical piece because at the heart of
15 this work is high quality learning and high
16 quality teachers. Those two things have got to go
17 hand-in-hand as we think about an assessment
18 system.

19 And I'm a state that 10 years ago we
20 tried to embark upon some of this and created a
21 local assessment system. And then Secretary Paige
22 said we could use that for our accountability

1 system and we buried our teachers. We need to do
2 this thoughtfully and make sure we build their
3 capacity to have the tools in order to move
4 forward.

5 You talked a lot today about curriculum
6 frameworks and is that critical. And I would tell
7 you it is critical. And using syllabi at the high
8 school I think is a very effective tool. Maine is
9 partnering with Dave Connolly right now and we
10 have expanded what they've been using in advanced
11 placement courses -- an online syllabi tool where
12 higher ed faculty review our high school teachers'
13 syllabi looking for alignment to the standards,
14 looking for rigor, higher order thinking skills.
15 There's tools that we can get at coherence and the
16 accountability system in, I think, some very
17 effective ways.

18 We also need to be very clear about what
19 are those exemplars. One of the things you'll see
20 as we go through the principles of this balanced
21 assessment system -- it's going to be critically
22 important for us not to be too narrow. Give us

1 flexibility. In our state we're moving toward a
2 standard space system, not a course credit system.
3 We're designing rubrics that will define all of
4 our eight content standards with Bob Marzano and
5 Dave Connolly. We want to share those as part of
6 this collaborative consortium in moving forward.

7 I'm going to go quicker. Some of the
8 things that we think are critical --

9 Our teachers need to be involved. It's
10 one of the best professional developments if
11 they're involved in helping to construct, helping
12 to score. I've had first-hand examples of that
13 with teachers in writing. We also need to make
14 sure that the focus is on teaching and learning
15 with these assessments. We want to make sure that
16 there are comprehensive learning progressions.
17 It's got to be explicitly clear. We need to be
18 transparent. This isn't about "I gotcha." This is
19 about raising our students' achievement and making
20 sure that we are world-class in all that we do.

21 We want to make sure that there are
22 multiple measures. You heard a lot about that

1 today. We don't have all the right answers. Give
2 us some flexibility. As Linda mentioned, several
3 consortiums are talking about how can we come
4 together. This balanced assessment, the Smarter
5 group who is looking at the technology and
6 adaptive assessments; the Mosaic Consortium is
7 looking at the formative assessment that you just
8 heard our NEA folks talk so well about. Those are
9 going to be critical to look at those new
10 technologies.

11 We want to create digital resources. As
12 a consortium, we can support one another as a
13 state and help to link and look at the
14 personalization of our students. Where are their
15 gaps? Those are things the consortium could work
16 together on. And we also think it's important to
17 build in a system for the moderation and the
18 auditing. And as I mentioned, I think this is an
19 incredible opportunity. We are very appreciative.
20 We are up to the challenge. The states want to do
21 this work. Thirty-eight have said a balanced
22 assessment approach. And so we look forward to

1 working with you and learning and giving us that
2 opportunity to scale this up for the country.

3 Thank you.

4 MS. WEISS: Thanks, Susan.

5 MS. MASON: Good afternoon. I so
6 believe I'm the last. I want to thank Joanne and
7 your team for all your hard work. This is the
8 third session I've attended and I'm amazed at how
9 the issues are being discussed, all the input
10 you're receiving.

11 I'm Christine Mason. I'm executive
12 director of the Center for Educational
13 Improvement. I've worked with charter schools in
14 D.C. In fact, I provided professional development
15 for probably 95 percent of the charter schools
16 over the last 4 years. I have a background as a
17 professor, a background in special education, and
18 I've conducted numerous research projects. I also
19 returned from four months in India where I helped
20 an international boarding school implement a
21 program for Cambridge accreditation.

22 With that I got to examine what happens

1 when you take a different approach to trying to
2 figure out how to get kids ready to graduate and
3 go on to college. And with that approach, some of
4 it I didn't like. I think the IB approach has
5 some strengths in regard -- in comparison to
6 Cambridge. But what I saw is opportunities for
7 the mid-course assessment where, in fact, in
8 December, we looked to see how well the kids were
9 doing. And then we're able to redesign some of
10 their learning for the second half of the year.
11 So I do think there is some value to that.

12 I have a few -- some of them
13 out-of-the-box -- suggestions. And some I think
14 reflect things that others have said. First, I
15 think technology is critical. We're doing a lot
16 of fun and innovative things with technology right
17 now. In my dream world what we do with assessment
18 would reflect that same spirit so that kids could
19 become as engaged with assessment as they are with
20 learning. So I hope that there will be room for
21 those sorts of innovations and we'll see
22 multimedia ways of using technology for students

1 to do problem solving. As they do that problem
2 solving, I think a way to assess their teamwork
3 and collaboration skills is essential. Some of
4 the other systems, such as Cambridge, do have ways
5 of looking at teamwork. And I believe that's one
6 of the critical skills for the youth for tomorrow.
7 So I hope that's included.

8 Research on human learning certainly
9 shows that people learn more and more quickly when
10 they're excited about their learning. When
11 they're engaged. When they're doing something
12 they love. And because of that, and I'm a strong
13 proponent of student self-determination, I hope
14 that we have some flexibility so that students who
15 show an interest in a particular area may be able
16 to elect that area or maybe two areas for more
17 in-depth instruction and rigor. And then perhaps
18 there is a threshold, perhaps at a lower level,
19 for a skills assessment in other areas. So I
20 believe those sorts of innovations are possible
21 right now and they might take some of the pressure
22 off of the teachers and students as we go forward.

1 In a similar vein, a number of years ago
2 I was part of a team that looked at exemplary
3 schools, secondary schools, for both students with
4 and without disabilities. And we looked at things
5 like career academies. And I think about those
6 career academies and what we might want in those
7 types of schools in terms of ways of examining
8 students' skills and strengths. Likewise, if you
9 look at some charter schools or other schools that
10 focus on the arts or technology or math or
11 science, maybe there should be a way so that those
12 schools would have a way to showcase what they
13 have targeted as their strengths. Again, I'm
14 advocating for flexibility in the approach as we
15 move forward.

16 Universal design for learning.
17 Wonderful concept. I'm glad to hear you're
18 thinking of that and that will be incorporated.
19 That gives a way for students with disabilities
20 and other students -- students who are English
21 language learners -- to participate with better
22 accommodations that are built into the test.

1 that brings us right at 5:00 to the last in our
2 perhaps seemingly too endless, but now concluded
3 series of panels that we've been running for the
4 past three months. We really appreciate your
5 participation.

6 I want to also thank all the parents who
7 came today from thousands of miles away for really
8 encouraging this work on behalf of your students
9 and your schools. I think it's an important voice
10 that we don't get to hear as often as we should.
11 Thank you for coming today and making your voices
12 heard. Thank you to everybody who has joined us
13 over the course of these different panels today
14 and for the last couple of months.

15 (Whereupon, at 5:00 p.m., the
16 PROCEEDINGS were adjourned.)

17 * * * * *

18

19

20

21

22

