UNITED STATES DEPARTMENT OF EDUCATION

+ + + + +

RACE TO THE TOP ASSESSMENT COMPETITION

+ + + + +

PUBLIC AND EXPERT INPUT MEETING
TECHNOLOGY & INNOVATION

+ + + + +

FRIDAY
NOVEMBER 13, 2009

+ + + + +

The meeting commenced in the Mystic B conference room of the Embassy Suites Hotel, 207 Porter Street, Boston, Massachusetts, at 9:00 a.m., Joanne Weiss, Director, presiding.

EXPERTS PRESENT:

EDYS QUELLMALZ
TONY ALPERT
EVA BAKER
RANDY BENNETT

STAFF PRESENT:

JOANNE WEISS
ANN WHALES
JUDY WURTZEL

TABLE OF CONTENTS

ITEM PAGE

Welcome/Introductions.....4

Expert Presentations.....17
Question 1

Round Table Discussion.....57

Expert Presentations.....86
Question 2 and Question 3

Round Table Discussion.....132

Public Speakers.....147

Conclusion

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

9:00 a.m.

MS. WEISS: I am going to do a quick run through of the stage setting that we did yesterday morning for everybody. So, I apologize to those of you who heard it once, but I'll try to do the fast version of it, so that those of you who weren't here, get the gist of what we're doing.

We're going to introduce ourselves in a minute, which we did not do a good job of yesterday morning. So, we will do that.

But I'm Joanne Weiss. I'm the Director of the Race to the Top Fund at the Department of Education and I just want to welcome all of you this morning and thank you so much for participating with us today.

The first thing that I wanted to make sure everybody understood is what event you're here for.

We did announce yesterday, the final regulations for the State Race to the Top Competition, the $4 billion competition

that is about rewarding states for comprehensive state-wide reforms.

This meeting is about the Race to the Top Assessment Competition and that is a $350 million set-aside that's designed to support consortia of states, who are implementing common standards, by helping to fund the true instantiation of those standards through the development of a comprehensive new generation of common assessments, and this morning's talk in particular is about technology and innovation.

I think you probably know from seeing the notices we put out, we have two day meetings in each of three cities around the country. The first day is a general discussion of assessment with different experts in each city. The second day, are deep-dives into particular topics that we think require a little more vetting and a little information, in order for us at the Department to do a good job of writing the notice.

And so, today, this morning's topic is technology and innovation. This afternoon, we're going to be talking about high schools.

Just to make sure everybody understands this, the applicants for this grant are consortia of states. Yesterday, the topic came up about why doesn't the Department just go straight to test developers and have them as the applicants. That's not how the law is set up. That's the first good reason we're not doing it that way.

But this is also a chance for states to really take ownership of this process and hopefully, do the right thing by kids, in ways that are perhaps, fundamentally different from what we're doing out there, although we certainly expect that they'll be a lot of partnerships that are formed with both public, private and non-profits organizations, in order to make this happen.

The time line is that we plan to have these conversations through the early part of December. December through February,

we'll be writing the notice. In March, the final notice will come out. So, this particular notice will not have a public comment period. Instead, we've designed this process at the front end, to help us get the right kind of input from experts and from the public, so that we design this notice properly, at the front end and looking at the process the we typically use.

We decided it wasn't the right process for this particular grant, that we really needed expertise in the Department, that we didn't have readily available and we also felt that it would benefit the state applicants, to be able to hear and share in these conversations.

So, we've designed, with the help of our colleagues at the Department, a very different process for how to develop these particular notices and that's what you're all part of right now.

The applications will be due -- the notice will be out then in March. The

applications will be due in June and grants will be awarded by the end of September, consistent with all of the recovery act funding that has to go out by the end of September.

The goals for this program at the highest level then are to support states in delivering more effective and instructionally useful assessments than perhaps we currently have, and in particular, we want more and accurate information about what students know and can do, in the area of achievement standards, how are students growing, how fast are students growing and are students, each individual student, are they on track to college and career readiness, so that by the time they graduate high school, they'll be ready for what's next.

We also want to make sure that these assessments reflect and reinforce and support good instructional practice, which we think is one of the biggest issues that we have with the common -- with the current set

of assessments, and so, it's sort of one little set of words up here, but it has a whole lot of baggage behind it.

The last one, of course, to make sure that we're thinking about and including English language learners and students with disabilities, not as an add-on at the back end, but as part of the design requirements and thoughtfulness at the front end, so that we design these right at the beginning and not try to tack on other purposes at the end, as an after-thought.

We want the data that comes out of these assessments to be useful to inform teaching, learning and program improvement, determinations on the part of the public, about whether their school is effective and whether their teachers and principals are effective and determinations on the part of parents and kids, about whether the education they've gotten, the trajectory they're on, is going to prepare them appropriately for what lies ahead for them.

We also have a few other requirements. We're in this funny place right now, where we are sort of the at the tail end of the current No Child Left Behind law, and we need to comply with that. At the same time, we know that we've got a design test that will take us across the transition and into whatever happens, as we re-authorize the elementary and secondary education act, going forward.

So, we're saying that at a minimum, we need to be able to develop tests in reading, language arts and math, and in grades 3 through 8 annually, and at the high school level. So, consistent with what we're doing now.

We're talking primarily about summative assessments, but by summative, we're not trying to imply that they are necessarily at the end of the year, nor that there's necessarily tests that happen only once during the year, nor that there is necessarily only one test in the assessments weeks that we

might be talking about.

So, we're talking at large, and it also doesn't mean that we're not talking about formative and interim assessments. It's just that we have to have summative assessments come out of this.

In addition, obviously, valid, reliable and fair is a sort of given for all of this, but also, that we're talking about tests that may replace the current test, as opposed to being additive.

The goals for these particular readings are that we're trying to do a number of different things by having these meetings and we've touched on this a little bit, but the first thing that we think is so important to do is paint a real vision for the country, about what assessment could and should look like, if it was meeting the goals that we have for it.

We've felt that we've got such a narrow version -- vision of what assessment is in this country, that we needed to have a

broad, public conversation with a number of experts who've thought a lot about this, both domestically and internationally, to just help us as a country, understand what assessment could do for our kids and for our teachers.

But second, we also needed to not be total pie in the sky. We need very concrete expertise, to help us in the Department, write the kind of notice that's going to elicit the kinds of assessments that we think are the right kinds of assessments, and that's going to be a tricky process. So, we will be asking people to be very concrete and put themselves in our shoes and say, "If you were writing this, what would you write."

Then the third thing is, we're also trying to make sure that we're given states, who are the applicants, as much access to high quality information as possible, so that when they write their proposals, those proposals will be as impactful as possible.

And so, for those of you who are wondering, the folks up here at the tables,

are our invited state guests. We've done outreach to all the states and invited them to come, and the folks in the back are public guests who have joined us.

This is the agenda for what we're going to be covering today. We have two different sets of questions that we've asked our respondents to take up, and we'll be talking more about those in a minute and each will have a sort of quick presentation that the experts will do, followed by a round table discussion.

All of you are going to be able to submit questions that you might have into our round table discussion. I think you got note cards when you came in.

If you've got questions that you would like us to pose to the experts, feel free to write them on the card and you can pass them to the people who are working here. I don't know, where did Anna go? Kevin is over here. Can you just wave, Kevin?

You can just pass them to our folks

and they'll get them to us or you can run out and bring them to the registration table and they'll come in and bring them to us, since we don't have breaks at the right times this morning. You're going to have to just sort of get those cards up to us and we'll take care of it.

I'm going to be a bit of a ruthless time keeper. So, everybody should pardon me, if I am a little bruttle, with making sure we have time to hear from everybody that we've asked to hear from.

We are transcribing all these sessions, and they'll be posted on our website, together with all the PowerPoint presentations that you see, as quickly as possible and I think you've got the information on how to submit written input, if you'd like to.

These are the different states that are attending today and we thank you all for traveling around the country to come be with us and hope that this is useful, and let me

just -- actually, I'm going to come back to this slide in a minute.

Let me tell you that the first question that we're going to pose to the panel is how they would recommend the different innovative technologies be deployed to create better assessments and why, complete with examples, and let me now just spend a minute going around the table and letting each person give a really quick, just name and title introduction, so that you know who all of us are up here. Tony?

MR. ALPERT: Tony Alpert. I'm the Director of Assessment and Accountability at the Oregon Department of Education.

MS. WEISS: Eva?

MS. BAKER: Yes, Eva Baker from UCLA, Director of CRESST, which is and R&D organization, related to assessment.

MR. BENNETT: Randy Bennett, Educational Testing Service, Research Division.

MS. QUELLMALZ: Edys Quellmalz,

Director of Technology Based Assessment and Learning Systems at West Ed.

MS. WEISS: And again, I'm Joanne Weiss, Director of Race to the Top Fund.

MS. WURTZEL: I'm Judy Wurtzel. I'm the Deputy Assistant Secretary for Planning, Evaluation and Policy Development at the US Department of Education.

MS. WHALEN: Ann Whalen from the Office of the Secretary in the US Department of Education.

MS. WEISS: So, terrific. With that, here is the first question, and the order is, Edys, I think you're first.

MS. QUELLMALZ: So, what's the --

MS. WEISS: He's got to put your slides up.

MS. QUELLMALZ: That would be more fun, if we used each other's slides, don't you think.

MS. WEISS: Sorry, this is what happens when you have a technology panel. Everybody is monkeying with all of their

technology until two seconds before the event starts and so --

MS. QUELLMALZ: That's not the right one. It's the one that says 11/13.

MS. WEISS: Are we there? Is this it?

MS. QUELLMALZ: It should be in color.

MS. WEISS: That's okay, let's just --

MS. QUELLMALZ: No, I don't think it's the presentation.

MS. WEISS: Okay.

MS. QUELLMALZ: Is that the one that's 11/13 that I just gave you?

MS. WEISS: Go to the next one. Yes, it's just, we've got --

MS. QUELLMALZ: Never mind, okay. It's a little odd, whatever.

Okay, so, today, I'm going to be sharing with you, some of the work and recommendations that are based on some projects that we're engaged in at West Ed and

it involves both the work of mine and my colleagues, Mike Timms and Barbara Buckley.

So, we thought we would start with an overview or our major recommendations and those are the following:

We think it's important in this competition, to not do things in incremental steps, but to try to move forward, in ways that are truly transformative.

So, breaking the mold, going beyond the functional uses of technology for things like easing delivery and scoring and reporting, to really take advantage of the capabilities of technology, and to do so in ways that are really going to help us better represent the kinds of thinking and reasoning in problems that students are being asked to do in the school subjects, and to use technology, partly as tools of the trade, so that if the kind of reading and writing that we are engaged in as adults are the kinds of technology tools for searching and finding and selecting and using and communicating.

Those kinds of tools are being slowly integrated into the schools and should also, in turn, be ones that we would see in our assessment tasks that can be technology based.

In particular, we think we should be focusing on what's not currently well tested in the static paper formats and limited item formats that we're currently using, and in particular, this relates to integrated knowledge, active processing and processes in general, as students go through reading mathematical problem solving, science and important problems in the domain.

Also, other questions today that we'll deal with later, about how we would reform the kinds of test form designs and the timing of the testing, how collaborative's can help off-set the costs and bring in multiple sources of expertise, and a way, we think, that it's very important is that there not just be aligned state and classroom standards, but there be common specifications and uses of

various kinds of testing banks, so that we can truly have some commonality of test designs, task designs and evidence from both ends of the spectrum and common platforms, meaning also, interoperable standards for authoring an administration.

We want to be sure people are aware of and take a look at the use of innovative tasks that are currently being used in large scale assessment. So, this isn't brand new. PISA, the international assessment in 2006, did a computer based science assessment, the 2008 has interactive computer tasks in it, again, focusing on the kind of inquiry that isn't well assessed in paper tests.

Minnesota has an online science test. PISA is going to be doing electronic text in 2009. NAP writing will be using word processors and editing and the 2012 NAP literacy framework, which will be entirely computer administered, will have scenario based items and lots of embedded tasks.

So, I apologize for the resolution

of this. I'm sure you'll have an opportunity to go back and look at the PowerPoint later on.

These are the advantages of technology that we can use, rich, authentic environments, multiple modalities, which we think are very important in both learning and testing.

Students can engage in iterative and active processing and multiple inquiries, multiple trials, lots of different ways of representing information, representing text, multiple response formats, so students have multiple ways of showing what they know and can do and letting them replay and iterate.

These multiple modalities, we think are particularly important to consider, in looking at our low performing readers, English learners and students with disability, and we have ways of dynamically representing various kinds of phenomena, particularly important in science and access to vast sources of information and data that again, can be cashed

and used in various ways in a summative assessment, but also, more broadly in classroom assessment, and we can support formal and informal forms of collaboration and assessment of it, which is a critical 21st century skill.

So, in work we're doing in science, we have come up with ways of characterizing the kinds of model building that students engage in, in any science system, and this is an example of how we have a generic model that moves across into particular topics like eco systems and then, how targets are associated with that, and again, you can look at that in more depth at some future point.

For formative assessment, we're very keen and have been using technology to log problem solving sequences, as students engage in iterative kinds of inquiry. We can give them immediate and individualized feedback, customized coaching and graduated coaching for them, and use the affordances of the technology, which for things that look

good and are very cool and are actually instructionally, very effective to cue and highlight and give students working examples, and also, we can provide adaptive tasks and items that are adaptive, in terms of how well students are able to deal with levels of complexity, which is sort of a different way of thinking about it.

We have examples in the PowerPoint that I recommend you take a look at, at some other point in time. This is a feedback slide, where students have gone through observing organisms and eco systems, identifying who is a consumer and who eats whom and what and what the flow of energy would be, in a drawing a food web, rather than reproducing a memorized food chain from their textbook, and this is a kind of feedback, where we're using highlighting and giving them feedback in one of the graduated levels.

In thinking about how some of the affordances we've been using in science apply to the new literacies, as they're being

called, digital or cyber literacy, there's an important report out of the National Science Foundation, about cyber literacy and what we mean by that, expanding the kinds, what do we mean by authentic tasks in English language arts.

We're talking about, of course, the range of discoursing some structures and what are tools of the trade in English language arts, that we want our students to come out of high school and be prepared to use in post-secondary and the workplace.

So, again, different kinds of modalities and images, graphic symbols, etcetera, and this allows students to engage in and search and find, comprehend, highlight, etcetera.

This is an example from one of our science tasks, where students write explanations. They see criteria for evaluating whether their explanation contains the appropriate kinds of points and then, evaluate their responses.

Similarly, in mathematics, you're saying, what are the important kinds of problem types? What are the tools of the trade in mathematics that go beyond calculators, and what are the kinds of tasks that are particularly important and valuable to assess with mathematics?

These are some from our science tasks, where students are doing analysis and multiple data representations.

So, we're looking for transforming assessment, going from what, where and how. Assessment is going to be done, again, talking about ongoing curriculum embedded benchmarks that have established technical quality, rather than on demand/on time, balancing the system, formative and summative, doing some kinds of matrix sampling and sort of a bi-directional flow of aggregating technical quality evidence from technology administered assessments in the classroom up, and sampling important kinds of well constructed tasks and using them in the classroom and looking at a

different set of arrangements of different test design models.

Look at that, 49 seconds. Is that good, or what?

So, also, I wanted to let people know that in this PowerPoint presentation at the very end, I was very sneaky. There is a set of appendices that are on the PowerPoint slides, that include a whole bunch of examples, if you're interested in looking at them, that I wouldn't have had time to go through here, and I also have some handouts of stuff, if people are interested.

MS. WEISS: Great, thanks. Let's --

MS. QUELLMALZ: Twenty-three seconds.

MS. WEISS: Yes. Rather than ask a question, we're going to move onto Tony and why don't you just re-introduce yourself for a second, Tony, and then you can start.

MR. ALPERT: Tony Alpert, Director of Assessment for the Oregon Department of Education.

So, thank you for the opportunity to speak with you about this topic. It's something that's very important to Oregon.

Oregon is -- has been heavily involved in technology and assessment for almost nine years now, and last year, over 99 percent of its students taking the general education assessment did so, online with an adaptive test, which as of January 2009, was approved for use as part of Oregon's AYP designations.

So, I want to talk a little bit about the benefits of an online adaptive test and then, talk about what criteria we might want to include in a RFP, to ensure that we have high quality technology systems that support assessment.

When you have an online adaptive test, you get a lot of benefits. One of the most important is student engagement.

So, rather than a typical fixed form test, where it's pre-defined which questions are presented to which students, the

computer dynamically selects those items, which are most appropriate for a student, based on the student's prior answers, and Gage Kingsbury talked a little bit about this yesterday, but I'm going to go into a little bit more detail.

But the benefit of that is that it provides tests that are more consistent with the student's abilities and typically, a student that's talented and gifted might be questions that are -- many questions that are too easy and a student who is lower performing, might get a lot of questions which are too hard.

With the adaptive test, all the questions are equally hard, relatively for every student.

An additional advantage to dynamically selecting items, it would be a much broader basis for evaluating the content standard.

So, with a fixed form test, you either have to spiral out a whole bunch of

different forms, or you have to be limited in the amount of content coverage that you're able to include on your single assessment, and with the adaptive test, we get to sample the content much broader spectrum than most fixed form tests are able to do.

We get more precise results for those students who are typically not well evaluated by a fixed form test, particularly those students who are extremely high performing and those students who are extremely low performing, and that's particularly important for students with disabilities and English language learner students.

The reason for that is that when the precision of a test, as many of you know, is related to the number of items, but also, the number of items that provide relevant information about the students abilities.

So, if you're selecting items that are at the students estimated ability, you're going to get more information about what that

student can do.

If you're giving an extremely test to a student that's talented and gifted, you don't really know what their true ability is.

But if you adapt the test and provide questions that are at the students ability, then you're going to get more information.

We report results immediately to the student. We -- and within 15 minutes, those results are available in an online reporting system to teachers. They can go in and see both their individual students results, the cumulative class results and even, the cumulative state results, all real-time, basically. It's a substantial advantage above paper and pencil.

So, when we're thinking about what's required to make all this happen, I tried to include some criteria. Much of this, Oregon already does. Some of it, we want to do with additional advances and as more states contribute to the -- our knowledge base about

adaptive assessments, what might we be able to do as a collaborative.

But one of the most important pieces of an online test is the secure connection, because we want to be able to ensure that the student is responding only to the test, with the allowable resources and not use the wealth of information that might otherwise be available, unless at some future point, you can track that and then, incorporate that into the scoring, but that's probably more of a broader vision than we might be able to cover within the realm of four years.

So, you want to have robust systems. You want to be operating system independent and operating system resilient. So, right now, when MacIntosh and Apple -- and Windows or Microsoft makes changes, that can have an impact on the state systems. So, you want a control for that.

You want to be efficient with a small footprint, one of the smallest possible

application size, with a minimum amount of overhead. You want it to be flexible, to function on a variety of network conditions, user friendly, to have a minimum of IT staff support, and of course, you want it to work with the systems that are available to districts.

So, you need to think about what would work on the oldest computers, with the least amount of memory, used among the districts with the least amount of resources, because if you build a system that the districts can't use, they're not going to be particularly happy.

So, with -- one of the key pieces that we need to think about with technology, and this is an issue with paper and pencil, but more so, when you're using online tests, it becomes more about an IT application, an information technology application, or equally about an information technology service application, as compared to a test.

And so, with a paper test, when

you're doing quality assurance, you're looking at the final product and you've evaluating, is it exactly what we want kids to see? With an online test, you're trying to anticipate, have we built systems, such that students are likely to see what we intend them to see, and intend for them to see, and that's quite a bit more difficult.

So, a project manager or multiple project managers is critical. We would want to an RFP, how the state or the consortia is going to attend to the system life cycle and development process, and I'm going to go through some of those key components quickly, but probably more people out in the audience that are more familiar with this process than I am.

But you want to have planning. You want to have specific tasks with roles and responsibilities. You want to have weekly project summaries, requirement documents and mock-ups.

One of the most important features

of online test, which is absolutely critical, because you don't get a second shot, once a kid takes a test online, the test is done, and so, the test plans are critical and documentation of both unit and systems tests, and then with a consortia, who is responsible for doing the user acceptance test, in both the development and pre-production environments is essential, and most importantly, you want to have strict change management policies with specific planned enhancement cycles.

So, with an online test, you can make changes very easily, but you don't necessarily want to make changes, just because you can, and so, having that controlled environment is -- with an established process, maintains the integrity f the assessment system, despite the fact that you are able to make changes whenever you want.

So, when thinking about additional criteria for the user interface, we want to think about the variety of computers that

might be used. Netbooks, we have a district in Oregon that's using those for Oregon state assessment. They're relatively cheap now, about $300 per student. But we can also think about notebooks, desktops, hand-held and tablets, and Oregon doesn't have a one-to-one computer ratio and I don't have the actual figures, but I think to the extent that we work towards a goal of a one-to-one student to computer ratio, we start expanding the possibilities for including instruction -- computers and instruction in the same way that we include computers and assessment.

We want to think about a variety of input devices. We want to think about speech to text, stylus, mouse pads, keyboards and touch screens, and we want to think about a variety of outputs. We want to think about text-to-braille, text-to-speech, text-to-ASL, even.

Then, most importantly, we want to think about a dynamic presentation of items that really looks -- that takes into

consideration, what the students abilities are.

So, if a student is visually impaired, then having an alternate item type, which is visually rich, may not be the best item that extracts the -- or identifies the mastery a student has of the content.

And so, we may need to swap out a different item that the student can access, but is associated with the same content that is otherwise being assessed by the innovative item type.

We have to think about what it takes at the district level to implement these assessments, and so, we need to think about IT support, and we provide, in Oregon, some diagnostics tools, but we have some room for improvement there and we need to think about how we can pinpoint problems that may occur at the desktop or at the server level or at the network level or even the internet service provider level, because really, honing in on where those problems exist is critical to

solving them.

We have added value if we can provide those solutions in pragmatic plain language.

We need to think about bandwidth as a resource, which includes traffic shaper's and there are cell phone wi-fi routers now that can solve some of the bandwidth problems that people may encounter, and we need to think about reports of capacity with longitudinal comparisons and we need to think about professional development for LEA or district IT staff.

So, I think that's my time, or do I have -- that's it?

MS. WEISS: Yes.

MR. ALPERT: Okay, I can take some of my other parts into question two.

MS. WEISS: Yes, we'll take them into our questions of you also.

MR. ALPERT: Okay.

MS. WEISS: Great, thank you. So, next up is Eva.

MS. BAKER: Good morning. Thank you for inviting me to participate in this activity. I'm going to premise my general presentation with some concerns about kids, and the idea that essentially, this is our opportunity, one more time. I guess we get it once every 20 years, I can't go back as far as you might think.

But at any rate, to really do something different, and I'm very concerned that we're not going to, all right. I'm concerned that -- I mean, I'm enamored of adaptive testing, but that's been going on since the early 80's, and education, you know, is sort of coming aboard.

We should be leap-frogging over some of these technologies, taking the pieces of them that make the best sense for us, but pushing quite ahead, far ahead, so that we're supporting what we want to support, and that is -- and you'll hear this throughout my 10 minutes, a concern for very careful design, so that the design of what it is you want kids to

learn and how that is to be measured, is the main thing that you're attending to, and that -- that design then becomes sustainable, if you invest in the right way, up front, and I don't think we've done that very well before.

So, I'm looking simultaneously and screen here, screen here, and screen here and I'm not very good at split attention.

Okay, so, what's the big challenges? Well, obviously, look at how many hyphenated words I have in the first sentence. We want multi-level, multi-purpose, open-architecture, open-source, scalable, top, down and bottom up, and by that, I mean that not everything comes from sort of the divine inspiration of states, down to the unwashed in the schools, but that there is a way for work and good ideas to move and down the system.

I think we want to spend our money up front on good design and by good design, I have a couple of notions.

Number one, while I like a lot that was in the description of the -- or the

expectations that the Department of Education put out in front of us, I would argue very much against separate systems for some of the formative and interim assessments.

I think they all have to be drawn from the same domain, or else we'll end up having happen, what happened before. That doesn't mean that all have to look alike, but they all have to be referenced back in a similar way.

I'll go on a little bit, explicitly about what I mean by the design of the `what' it is we want, not so much the `how', even though the first question is about `how' we would be looking at assessment differently, and much of what Edys has said, I agree with. So, I'm not going to necessarily repeat her comments.

Okay, what are our big challenges? In addition, when you define ways, and I'll suggest some options, for collecting validity evidence about what is going on, before we go into wide spread implementation, and that

necessarily hasn't occurred.

I mean, it's very hard to find the kind of validity evidence that one might want, for instance, that you have evidence that the assessments actually respond to instruction and are not simply measures of pre-existing traits of students.

We want to exploit new options. We want to do them in a shareable environment, so, interoperability, as Edys said. We want to do bells and whistles and `gee-wiz', and all of that, but only if it supports the kinds of multi-steps complex performance kids have. It isn't just, you know, trying to pick up web tools and apply them.

I believe we can save money in all of this, by designing, using authoring tools or using re-usable templates and objects, so that it's easy to make more examples of the kinds of assessments that you use, and that there's some nice-ites about how square criteria we're getting, which is probably too much down the needs for this, and training is

important for everybody who is going to be involved and I think what we're looking for is some mid-point in our assessments from just nice page turners on the computer, to dazzle.

So, here are some criteria for innovative tasks, other than efficiency.

Number one, what are the priority attributes? I don't think this my new one. We made some changes today, so, let's see if I can -- I'll just do what we have. Stay where you are and we'll just live with this.

So, what we want, obviously, are criteria such that creation of these things is confidential. You're not making one thing and then validating it and doing what you're doing and then having to start over every time. That's what drives costs up.

There is a picture someplace, that -- I guess I want to go back and see if we skipped it. Okay, there is a picture -- not that one either.

Essentially, what we've been talking about, I think Scott may have said the

word yesterday, about ontology. Ontology is simply a kind of representation of content that shows relations and, you know, is ripped off from computer science, medicine, engineering, it's a well-used kind of notation approach. It is used to make systems run.

And so, part of our -- a lot of our work is like taking from artificial intelligence, the parts of the conception that you can use in assessment, but that doesn't bring along all the costs associated with it.

So, everything gets referenced to an ontology or some representation that shows the relationships. This is not only a design tool, but a tool for professional development as well and has had some very interesting results.

You break that down to smaller levels, depending upon whether you're teaching a course or a unit, and I can show you more about that.

This is a kind of vision that we've provided for teachers, because they got -- it

was easier for them to take a look at that.

The other thing -- yes, this is -- okay, the other aspect of what needs to be addressed is the content of -- or skill set that goes with a particular content domain, and these are, you know, sometimes called cognitive readiness skills in the Military or in business, they are called 21st century skills by everybody and they all have -- everybody has their own set of them.

The one I'll bring to mind is adaptive problem solving and I'll show you, I guess later, an ontology for that.

I want to go back and show you some items types, since that's what we're suppose to be doing.

The kinds of things that I decided to show on test items, were not -- were relatively short, rather than things that took forever to go through, just because I thought that would be an easier start for what was going on in the classroom.

This is something you have -- you

know, realness of data, about essentially, people use a mapping function on -- when the kids are given a problem and instead of writing an essay, they write what they think are the key elements of it. There are -- the labels are directional and having meaning, so you -- the kids get scored not only on the propositions, but on the relevance of the links.

The scoring is automatic. These are not -- okay, so, this is a picture of how students would be doing a pull-down menu. This is something that feels like it's completely open-ended, but really isn't, and so, that you can have -- I mean, the number of choices is very, very large, actually.

But you can score it instantly, using expert, or various expert performances. We've used it from very little kids, up to college classes and it's especially useful for students with English difficulty.

This is -- people talk about the elephant that's in the room. Here it is,

although I'm not sure you can see it. This is a hand-held practice notion, showing how many elephants will it take to sort of stop of a flood and it's weight versus pressure. You can get the story.

This is one that sort of scares me, but it's also about unit and on the left are coils and the notion of finding equivalent units that you have to put together, to put out of you trampling's to progress, and this is something for, like formative assessment kind of work.

So, again, I would argue that you focus on these first elements, that is, what are the cognitive demands, what are the content elements you want, what are the linguistic requirements up front, that is, don't do it post-hold, don't do alignment post-hold.

Essentially, the picture that I presented of the database is one that should, at least that graphic, should rotated 90 degrees and you have to think of it as a top

of a database, under which there are items, instruction, professional development and other things, and I think I have run out of time and I'll have to re-order my slides next time.

MR. BENNETT: I have focused my question one recommendations on relatively high level principles, for how to deploy new technologies for assessment.

I'll get a little more specific in question two and question three, but detail really requires clarity of assessment purpose, because we should be driving decisions about what technologies to use, based on what it is we want to achieve in measurement, and we really don't have a lot of detail about assessment purpose, at this point.

My first recommendation is to start within the long term vision, five to 10 years out for next generation assessment system, and only then, work backward to a set of steps to get there, including significant near term ones.

Why? Because it takes two to three years to create review, field test, calibrate and administer a new parallel form of a paper and pencil multiple choice test.

If three to four years is the end state time frame for creating a next generation technology based system, the chances of achieving fundamental change are not going to be all that great.

The second recommendation is that in that long term vision, and to the extent possible, in the incremental steps, focus on such critical ideas, as using -- critical ideals, as using new technology to measure important competencies that can't be measured well, in the convention form, to help teachers and students adjust instruction and learning, to model effective teaching and learning practices, so that the assessment becomes a worthwhile learning experience in and of itself, and to make assessment fairer for all students, including those who are English language learners and those who are students

with disabilities.

Why focus on ideals? Because the danger is that if we don't, worthwhile, near term efficiency targets, like improving score turnaround, may dominate to the detriment of more fundamental goals, like measuring what's important.

I want to say a bit more about each of those sub-bullets that you may not be able to see, but that are labeled A through D.

Use technology to measure important competencies that can't measured well in conventional form. For example, having students use simulations of dynamic systems to interpret evidence, to discover relations, to infer causes, to pose solutions.

Having students mathematically model problem situations, using a spreadsheet. Having students write on computer and read on the internet. Reading on the internet is often non-linear. It's not the same as reading print.

Having students search for and

critical evaluate information on the internet. Having students respond to reading or writing problems that require the integration of many sources of text, of many different document types, not only print.

Having student fluently execute basic procedures, which are very important to know about for formative purposes. Having students carry out complex extended projects and having students assembly digital portfolios of their accomplishments.

Use technology to help students and teachers adjust instruction and learning. For example, when students summative test performance suggests the presence of either an overall proficiency deficit or a specific deficit, at the very least, provide teachers and students with formative hypothesis that point them towards skill areas of need, upon which the teacher and student should follow up, or we might route the student to a targeted diagnostic test that can do something similar.

Use technology to model effective teaching and learning practice. For example, build into test questions, tools that practitioners use, and that students should be using routinely in the course of their domain practice.

For example, planning tools for writing, graphical organizers and tables for representing complex text, concept maps, for representing physical or semantic relationships.

Use technology to make assessment fairer for all students, including those with disabilities and English language learners.

For example, vocabulary links for difficult words, where vocabulary knowledge is not what's being tested. Alternate representations of the same information, so text-to-speech or described graphics for students with visual impairments.

Alternate questions measuring similar skills at similar difficulty levels, when a class of question is important, but not

suitable for some students.

Recommendation three is to understand the benefits and limitations of each technology before deploying. For example, automated scoring is operationally fastest and cheaper than human scoring and sometimes, able to provide feedback on instructionally actionable performance components.

But in many cases, it uses limited proxy measures to predict a human score and practicing the proxies may lead to higher machine scores, but not necessarily to greater learning.

Similarly, adaptive testing measures with precision throughout the skill range, but in current implementation, measures only a sub-set of what it's important to test, potentially having the same unwelcome effects on instruction as current multiple choice assessments are said to have.

Therefore, recognize that there are trade-offs associated with new technology that

are best made by informed choice, not by accident.

Recommendation four is to manage risk. Most successful transitions from paper to computer delivery, have put substantial time into planning and many have used the phase approached implementation.

Oregon and Virginia are great examples. Both of those states deliver something on the order of a million and a half tests electronically each year, but it took almost a decade to get there.

Why is management risk especially important in this context? Because moving a large scale testing program to computer is a very complex, undertaking, requiring among many other things, hardware and software and compatibility in schools, extensive LEA training and student familiarization. Getting the appropriate infrastructure and knowledge into place takes time and planning.

Recommendation five, in the world of innovation, failure is a fact of life, but

it's a fact of life that can be put to good use. So, plan to fail, but plan to fail early, often, small and gracefully. Why?

MS. WEISS: That is not in the RFP.

MR. BENNETT: Why? Because we'll either learn relatively quickly, that an approach is unworkable or in the best case, we'll successively approximate over time, a practical assessment system with the least cost and harm to all concerned.

My last recommendation is to fund multiple consortia, so that significantly different assessment models and uses of technology can be explored and compared to one another.

I'd also suggest consider giving preference to models that already have an existing theoretical base and have been piloted.

Why? Because we know a lot about how to create innovative technology based assessments, including ones that ought to have positive impact on learning. So, we should

build on that existing knowledge.

However, we know a lot less about how to create such systems in ways that are affordable, practical, technically defensible, accessible and fair to all students. So, there's great value in funding multiple approaches.

So, the six recommendations are start with a long term vision, focus on critical ideals, understand the benefits and limitations of each technology before deploying, manage risk, plan to fail and fund multiple consortia.

MS. WEISS: Great, thank you so much. Let me point out that the agenda that I put up this morning, in my slide, is slightly different from the one in your program, just keeping with the fact that everybody is changing everything on you this morning. No, it's me too.

And so, what we're doing is two round table sessions instead of three. We thought it would give us time for deeper

conversation.

So, this is the first round table, around the first set of questions and we have a lot of questions from the audience that I think are about specific elements and uses of technology. So, I would expect at lunch that you might be mauled by the audience, by all kinds of questions they have.

But let me start by just asking sort of a big picture question, around the feasibility of doing this and some of the competing things that I've heard.

So, from Tony, I heard a question about -- or a recommendation rather, that we need to build to what's in to the technology that's there in schools today.

From Randy, I heard that we need to think long term and back-map from there. If we're planning an assessment that is designed to be delivered three to four -- three to five years from now, what's the design spec we're trying to hit?

We certainly don't want to put out

a test that's obsolete and the minute we put it out, because we designed it to technology and it's old, but nor can we afford as a country, to get too far ahead of what districts can really afford. So, how do we deal with that in the specs that we're trying to put out?

MS. BAKER: I think in the specs, you have to provide incentives for people to solve your problems for you, rather than to solve them in the RFP for them, and I think that you have to have -- I think I would lean toward Randy's vision, that is, we've run -- I don't think that this is the time, rather be safe than sorry.

I think that if we play it safe, we're going to be back where we were, with only a few different bells and whistles and people doing things on computer, online.

So, I think that what we really ought to be doing is identifying the functionality, that is, what it is you want these systems to do and the best vision that

you have of them right now, knowing that they're going to change and knowing that you're probably going to fail some, and trying to design to that, those notions.

There are proofs of concept, of those things, all over the place, probably as many things as you can think of, there are examples that have worked, and so, I think that you're not taking a step in the dark exactly. You can't. But just doing what we've done before, I think is not likely to be a successful strategy for us.

MR. ALPERT: So, I think Randy's point about phase-in and multiple consortia is really important in this.

So, if we allow the states and districts to use the assessment in the way that works for them and as Eva said, create incentives for them to be more aggressive in their adoption of technology, then I think we have a plan that is workable.

The danger is the piloting of this system. So, if you have a technology that's

so far advanced that most of the districts can't participate in the piloting, you're not going to have a good estimate of what its final impact is going to be.

And so, treading somewhat carefully is important, but there are states and districts that can participate fully in the most advanced technology and we should afford them that opportunity, and so, I think it was Scott, yesterday, who was talking about perhaps thinking less rigidly about comparability and more about flexibility and I think as we loosen the reigns on comparability, more of those opportunities become available.

MS. QUELLMALZ: Yes, I think that we have a lot of opportunity with the kinds of equipment, that currently resides in a large number of the schools.

I don't think we've begun to scratch the surface of what kinds of innovative tasks for both formative and summative purpose, let alone instruction,

could be done with existing technology that's in the schools now, and the idea of having a long term that goes out a decade, where you're planning on what kinds of things could most of the schools be working with in three years or five years, as a formative and summative assessment scheme, I think is very doable right now, and the idea of starting with small pilots with some of your districts, who are first year early innovators and the ones that represent a range of infrastructure and kinds of organizations, and then going to scale, is very important because you want to fail early, often and in a very small way, before you try to scale up.

I think there is -- one of the real dangers that needs to be avoided, like the plague, is comparability and equivalence. Those are more psychometric concerns than they -- if you're aiming toward common standards, that's the comparability that you're looking for, that you're aligned and you're testing things in a valid and reliable way with common

standards.

So, trying to make them comparable to your paper based tests, if you're doing innovative tests, is the kiss of death and it must not be considered as part of baby steps at all or you'll, I think, really limit what we can be doing, if we want to truly move ahead with what's right now, with the many kinds of examples that are in existence and have really gone beyond the small scale feasibility testing.

MS. BAKER: Let me say something about comparability because I'm getting the shakes.

Number one, I think comparability -- all my friends, except Ed, between different media, paper versus technology, as not very interesting anymore.

But I do think there's a comparability problem in the assessments that is going to take some invention, some thinking on the part of people who are in measurement or in learning, and that is that because it's

tied very much to validity and fairness, okay.

So, one needs to think about comparable tasks and elements within tasks, that can be identified, that go across different kinds of tasks, and my picture was one way of showing how that might happen in mathematics. Even though the task might look quite different, you could reference their performance to something that was going on in the original design.

But unless we think of new ways of dealing with comparability or in other terms, some reliability and consistency, we're going to be potentially putting kids in a situation where we don't know how fair the items are to them.

I think there are some of us working on this problem. It is not an easy problem, but it's something that has to be addressed. It was probably one of the reasons that some of the performance assessment activities, the last time or the time before, didn't survive, because we were not able to

show the cross-walks between different tasks and to explain really, what the domain was that the kids were competent in.

That, I think is easier to do in a technological environment, but shouldn't -- is something that shouldn't be skipped, and that's what I meant by comparability. I don't know if that's what you were talking about.

MS. QUELLMALZ: I think we need to think about comparable -- what's comparable with what.

MS. BAKER: Right.

MS. QUELLMALZ: When you define what you're looking at comparability for. My concern is that it not be comparable with a paper based test.

MS. BAKER: Right.

MS. QUELLMALZ: That's the idea, is to move --

MS. WEISS: Yes, Randy?

MS. BAKER: -- and supplement.

MR. BENNETT: Yes, the issue of comparability is one of, to me, commonality of

score or whatever characterization comes from an assessment, commonality of meaning from one student or group to the next.

So, you want to be able to be able - you want to be able to interpret the meaning of a score or characterization in the same way, regardless of what student it's coming from, what computer platform it was taken on, or whether it was taken on paper or on computer.

So, as long as you care about having that commonality of interpretation, then that concept is important. There's a tension between innovation and that commonality of meaning, because as we introduce new innovation, into testing, by they new types of tasks or new forms of assessment, we're changing the assessment and the scores or other characterizations may no longer mean what they did before.

If you're interesting in measuring growth, for example, that's going to be a problem, and that's something we have -- that

tension is something that's got to be worked out.

MS. WEISS: So, what I heard, as part of the answer, is don't specify things around technology. Specify what we need and what we're looking for and let people tell us how they would accomplish that.

So, that sort of leads into the question that you were heading toward, about how to develop an RFP that would -- walks that line that I think you put really eloquently in your slide, Randy, about enabling innovation and supporting innovation through a process like the one that we're about to go through, but at the same time, not creating giant failures.

So, if you were -- us trying to write a notice -- or if you're a state responding, how do you think about the role of innovation within a project like this? Is it something you think we should sequester pockets of money for? Is it something that you think states should put as a piece of it?

How would you advise us or states on that?

MR. BENNETT: Well, you certainly need some mechanism, because what you don't want to have is to institutionalize a next generation that becomes static and starts in its own way, to hold back student learning and progress in the same way that we believe the current system is and has for too long a period.

So, I think you have to build in an mechanism that allows innovation to happen on the side, keeping in mind the tension between innovation and commonality of meaning over time.

I think one way to do that is to think about the idea of chartering, and maybe that's not the right word, but chartering some set of schools in each consortium, that is going to move further ahead and try out new things that even the rest of the schools in the consortium are not yet doing.

Being sort of the cutting edge schools, that are given waivers from whatever

the current accountability system is, because you trust them enough and allowing them to pilot highly innovative ideas, a fair number of which are going to fail, but some of which are going to succeed and provide the basis for improvements and the next, next generation.

MS. WEISS: Other ideas that --

MS. QUELLMALZ: Well, I think this idea of pockets of innovation pilots is very important because if you place it in a high stakes environment, you're immediately going to stultify what can be done, and I think if you have some call for -- if you want innovation of a transformative sort to be part of what you're procuring, then having the consortia explain how -- and they may have multiple ways of suggesting this can be done, but it's not going to be for all districts, all schools in the next two years, or you won't be getting the kind of transformative innovation you want.

And so, one of the ways that many of us who have been working on this have been

doing is working with pilot districts and pilot schools and pilot states, to try things and having consortia look at that and say, "Does this have sufficient technical quality? Is this the kind of thing that we could see eventually contributing to our state assessment system evidence base."

MS. WEISS: And I think we're good at those small pilots and islands of excellence. I guess I'm wondering how we build -- I mean, I think preparing innovation was a good way to put the problem that we're faced being here, because those two words don't usually go together.

So, what is it that we do, to not only have islands of excellence where individual pilot schools are really pushing the envelope, but that we figure out how to take that innovation and when it is not failing, when it's good, bring it into the main stream.

MS. BAKER: Well, that's the -- excuse me, I would just say two things.

Number one, I wouldn't necessarily decide that all innovation gets sort of ghetto-ized in certain kinds of --

MS. WEISS: That was my concern.

MS. BAKER: -- in certain places, and then everybody --

MS. WEISS: That's exactly the question I'm asking.

MS. BAKER: -- else gets to do what they used to do. That, I think that there is --

MS. WEISS: No, I don't think it's used to do.

MS. BAKER: No.

MS. WEISS: I think it's the new thing and the --

MS. BAKER: I think there should be -- yes, there should be some technology things that we could sort of place our bets on, that we think that are sufficiently well known, if not in education, in other domains that could be easily moved into education, number one.

Number two, if you're going to have

cutting edge early adopter places, as -- which does follow an R&D model, which is what we've been doing for our whole lives, the real issue is number one, not doing them as if they were pilots, but doing them, so that early on, the first two things you're thinking about, other than the goodness for kids, is scalability and cost, so that they're not a sandbox for interesting, neat ideas, but that they are -- they have an end in view and that --

But the Government, I believe, could do is not only enable those, but also provide incentives for transition. The Navy, US Navy has a whole policy with our research for ONR, on transition agreements and that when you make something, you have to have a transition agreement and transition sites and a scalable story and a bunch of things that have to happen, for the work to go forward.

And while I'm not suggesting that's the right exact model, there are elements of it that might be helpful for you to think about, because otherwise, we could just be in

a loop of, "Oh, is it this fund, and look what I did," and not cracking through the bureaucracy, which, for very many good reasons, tends to be more conservative and not want to create, you know, difficulties with the public and so on.

So, I think that that set of incentives has to be very carefully scheduled, and with clarity, about what you mean for broader implementation at one point and scale up and what the cost constraint issue is, and that's really on the innovators screen right off, rather than innovative ideas --

MS. WEISS: Right.

MS. BAKER: -- as the first thing.

MS. WEISS: Tony?

MR. ALPERT: I think this is where the project plan becomes absolutely critical.

So, with clear phase-in steps, aligned in advance and probably more importantly, establishing clear criteria for success.

So, what does it mean for the --

for an innovative item to be successful? What are the attributes that we're looking for, and not that you should establish that, but that the respondents would establish their criteria and that would be the basis for evaluating to what extent do they move into the next phase, and then, I think you can some assurance that there is a solid intention behind the proposal, to scale up.

MS. WEISS: So, let me switch gears a little bit and say that yesterday in our conversations with the general assessment experts, we talked a bit about the tensions between looking at -- trying to measure student achievement against standards, student growth, teacher effectiveness.

Are there things that a technology based system could do, to make these different sort of measurements come together in a happier, more efficient way, or are these just things that we have to wrestle with, period, and decide what the priorities really are, and that's it? Like, is there something the

technology can do, to help us with the problem?

MS. BAKER: Well, in my view, the answer is both of those, both things, that there are some things that I could describe and people around the table could describe, that technology could energize and enable, but there are also priorities and as I said earlier, I think any time you're making a decision, and this is a tough one for people in policy and political environments, I think it's important to argue strongly for the quality of what the students are learning and able to do, as a principle role of what your system is for.

And that if there are secondary goals, such as looking at the impact of various treatments or teachers or value added in a variety of ways, not just, you know, classroom teachers, I think that's a secondary goal, to some people, it seems to be most important.

But I think it really is about

getting kids ready for the end point of school, whether that means they're going into career placements, Military placements, community college, four year colleges.

I think that whole range ought to be re-articulated because a lot of people have heard it to mean only four year colleges and I don't think that that's what is fully intended, or it may be.

MS. WEISS: Right.

MS. BAKER: The technology allows you to do -- to keep track of things in a much better way than we could before, particularly with individual student data sets and it allows you to focus in a way on individual children and their growth or their sets of experiences, both within classes or courses, and between, and we haven't been able to do that very well before.

We've been focused on, appropriately, I think on different sub-groups and trying to get parity in the system. But I think that breaking away from our sort of

group thing and start looking at each child as the unit, is really something that technology could allow us to do. It's a little scarier, because it has -- it multiplies the options a lot.

But I think technology also provides many more options for ways of addressing individual kids needs.

So, I think that there are some -- the thing that scares me most about value added and some of the growth models that have been proposed so far, is the extent to which they rely on one and only one form of statistical analysis, and I think -- and one and only one form of establishing evidence for validity, and I think what you could do is break open that, and have people propose things that are credible, but not necessarily in the same psychometric box as things have been in the past.

That doesn't mean all psychometrics is bad, it just means that there may need to be an expansion of the way we look at it,

otherwise, I fear that those ancillary ideas, value added, for instance, will drive people to not so interesting kinds of measures of student performance and therefore, less whole and important interpretations of the standards that people are trying to push forward.

MS. WEISS: So, yesterday, I think it was Laurie Weiss, who made a suggestion that we think about how to combine different kinds of testing models in new ways, and I'm wondering if -- so, I hear what you're saying and understand that.

I wonder if there's also different ways of combining these things and I think Laurie's idea was, do you use computer adaptive testing to get a quick placement along the learning trajectory of where a student is, at different points in time throughout the year, so that you can be monitoring and tracking growth, but when you find out where they are, you then dive deep in that area by giving them performance tasks and other kinds of assessments that are

appropriate to that kid, based on what your diagnostic just showed.

And so, you sort of now go off computer adaptive testing into performance tasks that the computer picked, based on the students.

So, are there things like that, that you can imagine, that would help us be more efficient in our testing and still very varied in the kinds of things that optimize for the different benefits of the different kinds of --

MS. QUELLMALZ: Well, I think one of the issues in that is again, definitional and linking it back to, what is it you want the students to be learning and how it is that the sets of tasks and items that you're considering adaptive, where they come from, how they've been designed and how they link to the ongoing instructional sequence of the students?

MS. WEISS: Right, yes, the idea was, it did. So, like, let's take as a given,

that it's not current, it's aligned with -- you know, that there is a set of standards that have some kind of order and that you --

MS. BAKER: Yes, so, if you design a different kind of adaptive testing, with tasks that are essentially within course, end of unit, aggregate, bottom up, instead of top down -- I worry about the top down.

If you take a summative test and chop it into little pieces and say, "This part is something that we could use. This component of -- to slice and dice and do adaptively," I think that will keep us where we are.

If there are new kinds of tasks and items that are developed, that go along with a progression of learning that goes within a grade level, in a domain that makes sense, and remember, you're changing topics all the time, then you would need careful thought about that as the overall assessment system, about what do we mean by growth across science topics, or across English language arts units or across

mathematics topics?

Then what segment of those could be given within reasonable points of time, when instruction has happened. Then it makes sense. But sometimes, people think adaptive testing is somehow just taking a chunk of what's already been developed at the summative level and plopping it in and that will set up back.

MS. WEISS: Randy, what?

MR. BENNETT: Yes, I guess I agree with Edys' comment. From my point of view of what was just described was kind of backwards, because I think I'd like to use the element or confidential items that one might have in an adaptive test for diagnosis, and the performance tasks to know what kids can do, because I think it's the performances --

MS. WEISS: That is what I was trying to say, I think.

MR. BENNETT: Yes, but it sounded the other way around.

MS. WEISS: I did?

MR. BENNETT: It sounded like, let's use the adaptive test to track growth and status with respect to the standards, and then, drop down into the performance tasks, to find out more about where a kid is having difficulty, and to me, it's kind of the other way around.

MS. WEISS: Right.

MR. BENNETT: Because I think the performance test ought to be the embodiment of most of the standards, not all. I mean, there are some things that elemental tasks are going to tell us about, more specific types of knowledge and conceptual understanding, that we can do more efficiently than with the performance tasks.

But I'd like the performance task to play a significant role in the accountability measurement, vis-a-vis the standards.

MR. ALPERT: But I think one of the things to keep in mind is universal design and whether or not those tasks are equally

accessible for students with disabilities and English language learners and the --

MS. WEISS: The performance tasks. So, the example that we've seen so far are very visually appealing and also, text laden and those become difficult to administer.

So, I'd also question -- performance task for performance task sake isn't really a good objective. So, what volume of additional information are we getting at the expense, potentially, of cost and accessibility, and I think those are critical considerations.

MS. BAKER: Yes, well, my answer to that would be, well, we would be getting notions of cognitive demands that can't easily be measured in selective environments.

The notion that students can take problems and implement various solutions to solve them, that are compositional, rather than analytic, that is, they have to put things together in new ways to make a solution, and it would energize the notion

that there are multiple paths that each student could take, to get to a sufficient level of competence and it's less, sort of, procrustean, in the way that everybody has to go down this same shoot, in order to get, you know, the check mark at the end.

MR. BENNETT: I think we could predict very well, how students could -- would perform on performance tasks from multiple choices tests.

So, if we're only interested in efficiency, then multiple choice test is going to serve the purpose, multiple choice adaptive test is going to serve the purpose.

But I think that the more important consideration is then, what is the wash-back that occurs, with respect to what kids model and teachers model in the classroom, and I think that's really where, you know, we're going to have to bite the bullet and say, "Yes, it's going to cost more. Yes, there are going to be issues that we're going to have to work out," vis-a-vis accessibility, we're

going to have to figure them out, because the alternative and the impact of the alternative on instruction and learning is just too great a price to pay, at least that's my view.

MS. WEISS: Great, thanks. Well, on that note, I'm going to move us on to question number two.

So, the second set of questions has to do with technology platforms. So, what's the underlying infrastructure that we need, in order to develop this next set of assessments?

This isn't necessarily only for innovative items. This could be for any type of assessment that we choose to develop, and just understanding better, what technology can do to help us with assessment development, assessment administration, scoring, reporting, that could both increase the quality and the cost effectiveness.

So, what's the functionality that such platforms would need, and then, how do you design these platforms, so that they serve multiple purposes and multiple masters, so, if

we design a platform that's really optimized for collaborative's to use in developing a collaborative, summative assessment, how do we make it also easily adaptable or useful to practitioners and professionals in the development of interim assessments or benchmark assessments, formative assessments maybe? Like, what are the other things that we need to think about, in terms of these platforms, as we're going forward, and then, any help that we can get on how to think about the cost and development of these.

So, with that, let's go in the same order that we did before. So, Edys, I think you're up.

MS. QUELLMALZ: Don't start that until we get to my slides. That could take us a while.

MS. WEISS: Sorry, we clearly designed a complex system here, for ourselves.

MS. QUELLMALZ: All right, question two. So, we're doing two, three, four in 10 minutes.

So, my colleague, Mike Timms, is in charge of the design of our infrastructure and design platforms, for the simulations we're working on in science and so, many of these recommendations come from his experience and also, our more general experience in using evidence centered design for thinking carefully about the relationship of what we want to test to the kinds of tasks that will elicit evidence of that and to the kinds of evidence that we'll be reporting.

So, in general, the recommendations are using design templates and re-usable components. We've all been saying that, so the more all of us say it, the more that becomes something to pay close attention to, in the development of technology based assessments.

We've been doing this in science for quite some time. These are life specifications shells and they're re-usable components that once you have some kinds of functionality, you can keep reusing these

objects.

You could reuse graphs. You could reuse lay-outs for doing experiments. You can reuse ways of helping students to search for and find information, and once you have those, you can affect a lot of efficiencies and economies.

In addition, what we've been finding in a project where we are working with six states on the use of simulation based science assessments, at multiple levels of their system, is we're using web-based platforms for easy access from any site and from browsers, so that we're not loading specialized software onto every single laptop.

This is an efficiency and a necessity for assessments like NAP, but within a -- hopefully, within some of these collaborative's, we can move toward using web-based kinds of ways of delivering.

Also, most of the scoring, he says, should be computer based. I added computer assisted, because I am a big fan of including,

of course, the teacher scoring and the student use of Rubrics, that I think is an important part of formative assessment, and of course, providing compatible formative and summative reports, both to students and to teachers.

We have here an example that you can look at, when you can see the slides in a slightly larger fashion. This is one that we've been using for supporting teacher scoring of constructed responses on our unit benchmark assessments, and teachers are given the kinds of prompts. They're given the Rubric, and then, they can scroll down all the student responses. This is a well-honed methodology that many people who are engaged in, in scoring student responses are quite familiar with.

The national language processing and automated scoring of short constructed responses hasn't been completely cracked yet. So, we are in -- in the paper, we note that there are a number of studies going on right now, with large scale assessments and short

constructed responses, but that they're not quite there yet, and when that comes along, the teachers can use those as a way of taking the sense of the class and still be able to delve in.

Mike has been thinking about the additional cost and benefits, as his technology budget starts overwhelming my design budget. Of increased site administration, costs are ones that initially, you deal with in the early pilots, because you need more skilled personnel.

They need to know how to go through and be sure that the machines are all working and that the -- everything is in order, in order to be running whatever the formative or benchmark or summative assessment is.

Item development costs are initially higher in the beginning and the initial system development costs are high, but in the end, you have the efficiencies of not printing and shipping, not scanning. This has no human scoring sessions.

I think in many places, the transition will be to some centralized re-scoring and double-checking, on up to human-less scoring and obviously, reporting can be much more efficient.

Our major interests, or at least mine is, in the reuse of these templates and components and building them into authoring schemes, and we've been doing this across multiple topics for our science assessment. It's eminently doable, once you've got the template, once you've got the simulation shells, once you have the story boards. For one eco system, it doesn't take a whole lot to figure out how to do it, for multiple new eco systems across the tundra and the Galapagos, etcetera, and similarly, these other kinds of replicable templates.

These, in turn, allow authoring by multiple developers and in turn, perhaps, by teachers. I know there are some controversy, about the extent to which teachers should be involved in certain kinds of authoring, with

professional development.

They're out -- they're used all along now, in terms of developing state-wide assessment, tasks and items, and in terms of formative assessment, they could learn how to use some of these as well, with some quality assurance checks added in.

The accommodations, I think, is an interesting issue and there are as many possibilities for how the technology based assessments, in fact, are reducing reading loads, when we have scenario based kinds of tasks and items.

You can convey a lot more about what it is students are being asked to do, what phenomena they're being asked to pay attention to, and you have multiple ways of representing that, that I think are harder to do in plain text.

So, there is trade-off there and we have a study running in the six states, actually, piloting in the three, where you're using some accommodations that can be -- can

be implemented in technology, where students can enlarge parts of the screen, where we have text-to-speech, and where students have various kinds of accommodations, as well as studying, with the ELL students and other students with disabilities, whether this combination of visual representations in fact, is helpful.

So, I think the empirical evidence is being gathered on the ways we can start turning to these new media.

Now, on question three, how can technology be used to develop high quality interim assessments? We've all heard the difference between the interim assessment and benchmark and formative assessment.

We all agree, or we better agree, that formative assessments are for use by teachers in classrooms, not to be graded. There's plenty of research about this.

So, in classrooms to be used formatively during instruction, to adjust instruction, they should be using these common

templates and specifications shells, the same kinds of tasks, but there are not -- the stakes attached to them, would pervert their use. So, don't do it.

These design templates and shells and story boards of reusable components allow us to develop pools of tasks and items that we can use for formative assessment tasks, for end of unit benchmark assessment tasks, that with the appropriate study, can be aggregated on up the system, into a state assessment system, and also, could be used to develop component tasks, as we did with performance assessments, that a state or a district could add onto the kind of summative assessment that it does.

So, we think this common core collection of secure and public tasks is a way to think about, and there was a paper that was written by me and Mark Moody, who had been assessment director in Maryland for many years, about how building common core collections of secure and public tasks allow

you to do some interesting kind of designs, test designs, and how you might embed secure tasks into end of unit benchmarks, etcetera.

So, that's a reference that's in the paper and in the slides.

So, this is the way it looks, which is where you take the benchmark assessments -- you take the benchmark assessments, you leave the formative alone and there, data can be scored and set up according to the proficiency levels and used either at the district or at the state level. Districts could use them and also aggregate on up.

So, this is a typical multi-level kind of design that we think is very promising and is -- and clearly, technology can help.

Embedded formative assessments, this is the way we use them in our science. It based simulation assessment work, where the embedded's go from online assessment to a progress report and the benchmarks go from online assessment without any feedback. Teachers score the constructive responses. We

use base-nets and then we have proficiency support -- reports.

These are the kinds of reports that go in the formative assessment, where you have students identified as needing help, making progress or on track, and then there are follow up reflection activities, whereas the summative class reports are ones that -- well, take my word for it.

And so, there are sets of recommendations that are taken from this national research council paper several years ago, from the committee on recommendations for state-wide science assessments, when science was just becoming required and this whole set of recommendations, again, talks about leveraging the power of collaboration, use specific plans and common specifications for connecting and integrating the assessments at the different parts of the system.

So, again, these recommendations are in the PowerPoint that you can take a look at later and the direct reference for them is

there.

MS. WEISS: Thank you. Tony?

MR. ALPERT: I guess I am done.

MS. QUELLMALZ: You're fast.

MS. WEISS: Good job.

MS. QUELLMALZ: See, technology can make it more efficient.

MR. ALPERT: So, I think a lot of what Edys already talked about, I don't need to go over. I'm obviously an advocate for computer based testing.

In terms of the efficiencies that we can achieve, online field testing is probably one of the most powerful. Gary Phillips gives a profound presentation on linking error and the dangers associated with it, which is, you create more linking error, to the extent that you spiral out and your unit of analysis is the school or the district.

You create embedded error that is reduced when you sample down at the student level and probably, Randy could probably go

into more detail about that than I.

But needless to say, you don't have to print out multiple forms, different forms, and achieve -- never the less, can achieve a great deal of efficiency.

You can also deliver multiple opportunities to students, without having to print and ship and sort. So, there are efficiencies at the state level, but there's also a great deal of efficiency at the school and district level, for districts that have to store huge freights of paper tests. They appreciate getting that room back, especially with over crowding.

Oregon, it's hard to estimate the savings at this point, since we've been online for such a long time. But at this point, and it's not really a fair estimate, because relatively few of our tests are administered on paper, but about $15 per test is what our paper test costs and about $4 per test are what our online tests costs, and I distinguish cost from expenditures, although I said cost,

I meant to say expenditures, because cost implies value.

And so, although our overall expenditures have not decreased, our value has increased dramatically, and so, what we've done is we've shifted the funds that would have otherwise been expended on a disposable product, like paper, and have incorporated more into building an infrastructure and a design that's sustainable, doesn't depreciate like paper does.

So, additional advances that we've had is really getting into the test development process. That's where we were still paper focused.

So, what we used to do until very recently, is we would do all of the test administration online, but much of our test development would have been on paper.

So, we'd do item writing on paper. We would shuffle the papers around and we'd enter those into the system. Now, we're going to the next level, which is creating the

entire system online.

So, distributed item writing, which I think is an amazing opportunity to incorporate teachers into the process, getting back at one of Joanne's questions, about how can we use technology to support practitioners and maybe increase buy-in more, which would reduce potentially, some of the problems associated with the use of the tests for evaluation, but incorporating more teachers into the process through distributed item writing can be particularly powerful.

Then, taking that next step and going through the online collaborative content review, to establish and to remove any bias and do the edits directly in the online system, rather than editing them on paper and data entering the corrections into the online system.

What we still do, or have done recently, is conduct the alignment review on paper. So, even though our items were store online, to have an external reviewer conduct

the alignment review, we printed all those items out and we shipped them off.

And so, getting to that next step of incorporating the review into the system itself would be particularly powerful, and then, of course, developing the test online, is infrastructure that can be used for a variety of purposes, which I'll get to in a second.

So, for online writing, Oregon does have a writing assessment, even though its reading and math assessments are entirely multiple choice. We do have an on-demand writing assessment and we're moving into making that more widely available to more students.

We are going to use artificial intelligence scoring, but we're going to look at, as hybrid option, where we take the artificially intelligent -- intelligently derived score, the computer based score, and compare that with a human score and then use a human to determine which one is more correct,

and that gets rid of some of the problems that we might have, if we were to use the artificial intelligent score, in and of itself.

We are going to combine paper and online scoring into a single system. So, at one -- soon, we will scan in any paper writing assessments that were written in paper and use the online scoring system to score those as well, and that has some substantial advantages.

We can do a random allocation of validity and qualification papers, so -- which will be blind to the raters. At this point, they have a pretty good idea of which of those papers are reliability and qualification papers, which reduces some of their effectiveness. We'll be able to do that in a more seamless manner.

We can have random back-reads, so that an expert will be able to review the quality of the ratings that are conducted by the scorers, but most importantly, we'll be

able to have real-time statistics, regarding reliability.

And so, if there is a problem in the scoring process, we'll be able to intervene immediately, rather than waiting for those results to come back to the scoring director.

So, how do we make this happen and be more efficient? As Eva said, we need to provide incentives for the LEA's to manage their resources well, but we also, more importantly, need to provide incentives for the state or the vendor, to be more efficient.

So, we would eat up the resources of the districts if they let us. So, by having -- by recognizing that they have restricted resources, we make our assessment more efficient by reducing the size of graphics, by reducing the size of the overhead, the amount of overhead that the browsers use and we are just very careful about using up their resources in the most efficient manner possible.

But it's important to recognize that bandwidth is not the barrier that people may perceive it to be and we haven't tried this imperially, but based on a calculation on the size of the files and the amount of data transmitted, in Oregon, we should be able to test about 12 kids simultaneously, using a 56k modem.

It's not that big. The bandwidth needs aren't that great, and so, we need to -- we shouldn't use that as a barrier to moving forward, necessarily, and most importantly, though, good planning and flexible testing windows can compensate for capacity.

So, Oregon has an eight month testing window. Districts can test any point during that eight month, any of their students and each student has up to three opportunities during that eight month window, and that flexibility allows them to use the system in the way that works for them, but it also recognizes that they have limited sources that they need to manage.

So, as we think about helping them be more effective, we need to think about assessment literacy. We need to think about training and certification on the use of the system, IT literacy, to the extent that the computer systems, the networks that the districts design are not the most effective. That's going to limit our capacity to help them use our system and probably one of the more critical aspects is having a multi-tiered help desk system, that can help districts resolve problems that do occur.

So, how might we incorporate this into a summative assessment? Edys talked a lot about how this might work, but I think from a design perspective, we could think of the formative assessment process as a gateway into the summative assessment process, where teachers are working on item writing for the formative assessment and then, perhaps they graduate into the summative assessment.

To the extent that the same tool set is available for them, for their formative

assessment, as it is for their summative assessment, that might be more effective as an opportunity.

To the extent that technology is more widely available in the classroom, I think we can imagine that the interim assessments could align better with the summative assessments, both in terms of the manner in which they're administered. If every kid had a computer accessible to them, then instruction -- formative assessment, instruction and the summative assessment all could better aligned and to the extent that that's practicable, it allows the instruction to be consistent with the students real world application and the manner in which students would be assessed.

MS. WEISS: Thanks, great. We will now move to Eva.

MS. BAKER: Just as a word of explanation, to the audience, what happened is that we have each made some revisions on our slides this morning and so, there's been some

disjunct in connecting up, and that's what's going on. We're not quite as addled as we may appear.

So, I have to I have to jump around a little, just to make sure, which -- because I have 75 versions, the poor guy is trying to make sense out of.

So, let me show you this picture again, which was a picture of an ontology in math, which is at a very high level, talking about concepts that -- or topics that could be course level or below.

Ultimately, when you get down to something that's like number theory, as used in algebra, you would then be referencing back pieces or entire items to this.

So, the notion of post-hope alignment is not so important, if you can document that the items have been generated from these kinds of things. It isn't a matter of somebody coming in and putting new eyes to it, necessarily.

I showed you a picture of cognitive

demands and I'm reviewing that because this also would be part of the development platform that you would use, that is, this is a set of resources that get operationalized, the way you create items or you create tasks and you review and vet those tasks.

So, my point is that to think about a platform, you have to think about the process that it's enabling and some of the processes have been described before, but let's say, number one, we want this representation of knowledge and cognition. That's the top level of a database. That's not just two dimensional pretty picture for you, okay.

So, rotate it, in your mind, all three of you who can do that -- I'm just joking, 90 degrees and it becomes like a top of a birthday cake, okay, and then underneath, in the birthday cake, are the items, the evidence about those items, necessary instructional routines, professional development, connections, and all of them are

anchored to these representations, which are drawn from the standards created by teachers and other experts. I mean, they're people and judgement based. They're not spit out of a computer.

Okay, so, if you buy into this idea, that some representation, not necessarily our's, is there, to provide not only explication of what the standards are, but shows explicit relationships or hypothesized relationships from them, and I would like to, as an oppositive, say, learning progressions are sort of a theory in some cases. They really need empirical validation and there are some strategies for doing that, in ways that don't take a whole year.

But that's -- when we talk about learning progressions, we shouldn't talk about them as if we know truly what they are. There may be many different ones and they may play out in different ways.

So, we create tasks, based on these models and from research on learning, and so,

our tasks are explicitly as -- not only universal design issues, but we were explicitly trying to reduce irrelevant cognitive load, that is memory, that kids need to keep in their heads, while they're solving the problem.

We want to use explanation, to produce deeper understanding. We want to provide a variety of situations for students, so that transfer potentially is enabled and they can infer schema or patterns that they can use in other settings.

As Edys has said, and not surprisingly, since we probably did this together, authoring systems and reusable task elements are really important. We have done them across subject matters, across grade levels. They get implemented into the subject matter and changed and adapted, but for -- in many ways, they're the single greatest mechanism for saving money that we have every had.

Similarly, we've had scoring guides

that we have made parsimonious, so that they only get adapted to particular subject matters in -- as needed, rather than starting over with a trait based scoring approach for every single prompt or every single problem. That just is -- was done before, when we did performance assessments. It's guaranteed way to spend money that you don't need to spend.

Last of all, if there are going to be items banks, which either are fully enabled or at least held -- or are put there, so people can populate them with pieces of our system, we want to make sure that there's a part of that, that's safe for transfer items, and that is for both the accountability side, as well as for practice in interim and formative settings.

You don't have items that kids have not seen before and that -- that instantiate the goals that you have. You'll be back to practicing the test, which is not what we really want to enable.

So, I'm going to go down through

all of this, back here.

Okay, I showed a picture of about cognitive demands, one of which was problem solving and this, I'm only showing you to say, or illustrate that the way we've done test specifications in the past has been fairly global and now, we're talking about things, if we want to make them work on a computer, and actually will want to know what we have, we have to spend more time in front end analysis.

These characteristics of problem solving were developed by some cognitive psychologist whose name I won't give you, because I haven't cleared it with them, and we're in the process.

They took -- they worked these things through, this is a group of four people from different institutions, and then they put it in our mapper, essentially, and made what they thought was a problem solving ontology.

We are doing cross-walks with that ontology and the math and literacy and science ones that we -- and history ones that we have

in place, so that one possible way of sorting through the database, for instance, is to look at attributes of problem solving as the child does that across different subject matters, as opposed to just simply attributes of subject matter.

So, the assessment development platform, it needs, in my view, also to focus o all the assessments that are going to be developed. There are just spaces that are set aside for various functions. If you're doing accountability assessments, you draw -- you know, you have permission to work on it and you draw from certain elements.

But obviously, there needs to be collaborative spaces for whatever user groups there are. There needs to be training available, both for the teacher based scoring, the notion of models, how you do criteria ratings, teacher scoring or raters, or whether you're using automated or semi-automated systems.

There could be, and I think it

wouldn't hurt, while you're designing this, to put in professional development for teachers and parents to assist in learning, when the kids performance, let's say in a formative setting, it's not optimal.

Our experience is that despite the wonderfulness of many of our teachers, that are some teachers who, when they find kids not doing well on something, don't have a full repertoire of what the next step could be, and so, I think what we want to do is expand that.

There could be a model instructional options and units. We would want an expanding set that is a growing accretion and -- but organized set of annotated examples of student work, in the areas that we're looking at. We want to talk about alternative automated approaches to scoring. There are -- you know, we have heard about AI based ones that are -- you know, may be caught -- some are cost intensive. There are some blended models. There are some very -- quite different models that I think, if the

Government is incensing something again, you know, on the list, ought to be the kinds of automated scoring systems that could be used well and not gained easily.

Then, guidance for feedback, including timing requirements and then there's a whole discussion of privacy protection for students and also, another one that we probably don't want to open up right now, but I'd be happy to sometime, about whether tests ought to be secure or not, but most of you think they should be, so, okay, then. I have a different view. I think they should be open.

Okay, so, here is another picture of the algebra thing broken down even to a more precise level. In the summative assessment space, again, the standards and representations collaborative space, another review component here of the design of linguistic features or other accommodations, evidence of the instructional sensitivity.

This is the most important validity

evidence I would purport, that we are using lots of assessments in -- to measure the effectiveness of schools, that have not been demonstrated to show that they respond to teaching. That is, they may be just showing us where kids are.

And so, my colleague, talking about correlations, makes me very anxious about what assessments correlate with other assessments, because I mean, if we're into correlations, we should just measure height and let go.

So, I really am very concerned that there is evidence that the assessments that are used grow, as a function of what students are taught, they're teachable and learnable and that seems obvious, but not -- the evidence isn't fully there, and then data collection and analysis services.

So, the metatags in this database could allow you to access things through the representations, what situations tasked, what kinds of students you have, criteria, linguistics, many, many different ways, sort

of sub-prism and you can look through it. I heard prism was a bad word yesterday, but presumably, not for this audience, and we're looking at evidence by student.

I think that you need to -- and this in the planning notion, to have accretion in improving the cycles, so that you can address the coherence, which I think is sort of a better term, than alignment, of those interim formative systems, with what it is you're intending to do.

We want validity data prior to implementation, cost targets and comparative benchmarks. Here is California on some --

MS. WEISS: Okay, Eva, we need -- I just don't want to take time from Randy.

MS. BAKER: Okay, stop? Right, oh, I do. Sorry, I didn't realize the numbers were going up.

But just to show you, there was a picture of that, that was asked for, and we can talk about, maybe in the question period, some things about cost and how to design to

it.

MS. WEISS: Thanks.

MS. BAKER: Sorry.

MR. BENNETT: Okay, I'm going to talk about question two and three, separately, in sequence.

So, question two focused on the platform functionality and my first recommendation comes in two parts, and the first part is that the platform should support the development presentation and scoring of assessments, that represent as fully as possible, not only the standards, but also, the results of cognitive scientific research.

The reason is that that research can help us translate those standards to test specifications, as well as to what should be going on, in terms of classroom practice, and I think this was part of Eva's point.

That research says we should be measuring higher order thinking skills, but we should also be measuring lower level components too, like declarative knowledge and

automaticity, the later of which requires the capability to collect timing data.

We should be measuring problem solving processes, which can be especially useful for formative purposes, and which requires the capability to collect keystroke and mouse click data, and we should be modeling the habits of one, the characteristic of proficient performers and their domain, which requires the capability to integrate tools and -- to integrate tools and performance criteria into test questions, so students learn to use those tools and internalize those criteria, in their own work.

To measure and model this array of competencies effectively, the platform should support the development presentation and scoring of assessments that include dynamic stimuli, like audio, video and animation, constructed responses of all types, simulations, information, resources and scenario based extended exercises, and to Tony's point, which I think is a very

important one, not simply for the sake of having them, because if we just use them for their own sake, we're not going to have the effect that we need to have.

We need to use them because they offer us either some measurement opportunity we wouldn't otherwise have, or some opportunity for modeling for teachers and students, something that we couldn't otherwise model.

But the platform should also support the development, presentation and scoring of traditional test questions. Why? Because the types of tasks encountered and competencies required in the workplace and in advanced academic settings can't be effectively represented through traditional types of tasks alone.

Recommendation two was that the platform should support frequent measurement with the capability to aggregate information over time, to form a summative judgement.

That frequent measurement could

take the form of multiple summative tests distributed across the school year. It could take the form of one or more standardized projects. It could take the form of electronic portfolios or some combination of those elements.

Why? Because we should be able to make more meaningful and fairer decisions about students, about teachers, about schools and education systems, if we combine evidence from multiple time points and from multiple sources.

Recommendation three is that the platform should minimize the influence of irrelevant factors on performance. That means including tutorials, practice tests, formative assessments and instructional exercises that use the same interfaces, representations and tools as found on the summative assessments, so, students become familiar with the mechanics and tools used in testing, and we should account for the needs of students with disabilities and English language learners, in

formatting those mechanics and tools.

Why? Because the test performance obviously, should depend only upon those aspects of student competency that are the intended targets of measurement.

Recommendation four, the platform should support an advanced type of adaptive testing. Traditional item level adaptive tests require short tasks that are machine score-able in real-time.

So, we should be looking toward new approaches, to adaptive testing. For example, a traditional adaptive test section that routes students to an appropriate extended constructive response section, that might not be automatically score-able, or a multi-stage adaptive test, in which each stage consists of an extended scenario based task, including both machine score-able and extended and constructed response questions, with routing from one stage to the next, based only on the machine score-able items.

Why? Because adaptive tests can

provide some important advantages, but in their current forms, they don't measure everything that's important to measure.

Recommendation five, the platform should support online scoring, online human scoring, as well as automated scoring and their combination. This was Tony's point also, which I thought was a very important one.

Online scoring allows for geographically distributed human rating with real-time monitoring of rater performance.

At the same time, significant advances have been made in the automated scoring of many different types of tasks, essays, short text responses, math equations, numerical and graphical responses and some types of oral responses. In addition, detailed performance feedback can often be provided.

Thus, these approaches can potentially make scoring cheaper, faster and better, especially when online human scoring

an automated scoring are used in combination.

My last recommendation, with respect to question two, is that the platform should make it easy to switch testing vendors. The platform should represent test questions and automated scoring models and common formats, so that questions and scoring models can be moved from one vendor system to the next, without undue time, cost and effort.

Why? Because states should be able to make vendor selections without having to bear the cost of repeatedly converting test content and scoring.

Okay, so, I'm going to move to the next set of slides and that one focuses on supporting interim assessments.

Recommendation one is that the platform should allow for the possibility of making interim assessment part of the summative system, as well as for incorporating other sources of evidence into that summative system, like projects and portfolios.

The interim assessments themselves

should be composed of the same mix of innovative and traditional tasks as found on the end of the year summative test. They should incorporate learning progressions, where those are available and they should be a learning experience in and of themselves, not just a test.

Why should the platform allow for this possibility? Making interim assessment part of the summative system would distribute the evidence used for summative judgement over additional sources, reducing the influence of a single end of year test and employing essentially, the same model that teachers now employ in awarding course grades.

They typically take some weight average of mid-terms, final, quizzes, classwork, homework and the like.

Note that the more interims or other sources of evidence there are, the less each counts individually. So, it's less likely that a student, a teacher, a school, a principal is going to be held to one

unrepresentative performance.

Making the interims part of the summative assessment would also more frequently model for teachers and students, the competencies and the tasks that are critical to proficient domain performance and the learning progressions that are likely to leave there.

Finally, to give timely, but preliminary formative feedback, pointing teachers to students in areas of need on which teachers ane students should follow up.

Recommendation two is that the platform should have the capacity to offer a variety of teacher optional curriculum embedded formative assessment materials, linked to the standards and to the summative assessments, as embodiments of those standards.

The formative materials should include traditional items, targeted at specific component skills, innovative tasks, targeted at skill integration, problem

solving, reasoning, critical thinking, conceptual understanding and the like, including scoring Rubric and exemplar student responses illustrating different score levels.

It should include pointers to additional relevant instructional materials, learning progressions, where they're available, linking those items, tasks and instructional resources to the standards and finally, guidelines for teachers on a suggestive process for using all of that stuff in a way that allows them to understand better, what students know and can do, at any given moment, so that they can effectively adjust instruction.

Why is it important for the platform to have formative assessment capability? Because on its own, interim assessment is insufficient for supporting classroom assessment needs. Teachers and students need curriculum relevant items, integrated tasks, Rubrics, interpretive materials, and instructional resources they

can use on a daily basis, if they're to focus and make progress toward achieving the standards.

Recommendation three, the technology platform should have the capacity for teachers to add, modify and share these formative materials. Teaching context and student populations vary. So, the ability to customize is important, but many context and populations are similar enough, that contributions by one teacher may be useful to other teachers, so, mechanisms for sharing are essential.

My last recommendation is that the technology platform should have the capacity for teachers and students to score constructive responses of all types, at least formatively.

That platform should be able to present Rubrics exemplar responses, qualification sets that help teachers determine if they're applying the Rubric accurately, tools for annotating responses and

recording scores.

Why? Because teachers and students can develop a shared understanding of what makes for good performance in a domain, through scoring, particularly if you identify the features and responses, that make those responses of higher or lower quality. I'll stop right there.

MS. WEISS: Thank you. That was terrific, and actually, you answered -- you just answered one of my first questions that I had for use I the roundtable.

Let me talk for just a minute, before we go into the roundtable, about what we're trying to do.

So, we are running quite a bit over. I'm going to ask that we do about only a 15 round table discussion. We'll take five minutes of break transition time, to move the podium up for the public speakers and we should still end about at 12:30 p.m., maybe a couple minutes late, for those of you trying to plan planes and other things.

So, let me ask a question about -- that you definitely started helping us answer, Randy, but maybe some of the bigger picture questions, about what we should ask the consortia to think about, as they're trying to figure out what types of technology platforms to support.

So, should -- what is it -- what does it mean to be open? How open should these be? How important is it to be so tightly coupled with the kinds of goals and purposes, that you become less flexible, because you're trying to accomplish specific goals that you have, versus being more flexible and more open to all people?

I heard you say that you need to be able to move items easily from one platform to another. What does that mean, as we're writing specs or responding to the note -- or the states are responding to the notices that we put out?

What kinds of specific concrete advice can you give us and the states, around

this question?

MR. BENNETT: Well, there are certainly existing standards for common -- for formats, content formats and Eva -- I don't know if she mentioned it in this session, but Eva mentioned one that the Department of Defense uses, called score-em, and there are -- there is also one that is used in educational software for test questions called QTI. I don't remember the meaning of the acronym.

But there are such standards and at the least, it would seem sensible to require consortia to implement content, so that's test question, scoring models and the like, in ways that can be easily transported, given that those test questions and scoring models are the intellectual property of the consortia and not of the vendors.

MS. WEISS: Do you think there is off-the-shelf answers here, Eva?

MS. BAKER: Well, I think there are -- I think they have a whole list, actually,

from my 50 minute presentation, on off-the-shelf technologies that are used in large scale in other places, and that I think there are ways we can do what we're doing.

I think the major issue the principle underlying what it is we're doing. Randy and I guess, all of us have been talking about those, and that is that we are trying to optimize clarity and flexibility, at the same time.

The notion that in your question to Randy was, when -- would we be inflexible, because we were trying to link things so much to the content that we had?

Obviously, that content is in motion as well, maybe not the top level standards, but the break-outs underneath are something subject to empirical change.

So, I think that what we want to do is have these platforms with the minimum amount of core information that we need, adopt the notion that there are places for people to add and that you can keep adding and vetting,

because just adding things from teachers, may not give you what you need.

So, that what we have to do is, have approaches for review, that are built in, and some of those can be automated and some of those can be not, but I think that will be the way to use the things that people already have in place, as design architecture's for very complex -- what's in Military systems, there are ways of moving that same kind of architecture into these systems, and that's what I think we need.

MS. QUELLMALZ: And I think too, that maybe part of the RFP or a criteria would be some kind of demonstration that -- some point in the design and feasibility stages of developing these new systems, where they -- each collaborative could demonstrate how it could take some of its architecture and tasks and items and they could be used by another collaborative, because you don't want collaborative silos either.

MS. WEISS: Right.

MS. QUELLMALZ: You know, so, if you -- maybe part of the RFP would be at some point in time, there should be some proof of concept demonstrations that you could run in anywhere in the US, that, you know, has the technology based platform and then that's, you know, sort of proof of the pudding, whether it's really inter-operable or not, and would certainly keep people from developing -- the content may be proprietary, but you don't want the infrastructure to be so proprietary that it can't be used for the greater good.

MR. ALPERT: But on the same token, we don't want to necessarily restrict the consortia of the collaborative to a pre-defined set of -- or architecture.

So, we would want to keep the innovation and then like you said, we want to, at some point, point to an -- or have an end stage or an intermediate stage, where there was some kind of collaboration to achieve a common format.

MS. QUELLMALZ: So, Tony, maybe

you're a good person to start out the answer to this question.

What advice would you have to states that are working in a consortium, to think about how to develop a platform? This is -- I'm asking sort of a project management question here, governance, leadership. How do you organize for success when you're trying to do something at this scale in technology, across the -- as cross states?

MR. ALPERT: Yes, it's -- cross state consortiums are hard, because you --

MS. QUELLMALZ: That's why I'm asking.

MR. ALPERT: Yes, everyone has really great ideas and they all want to get them out on the table, and so, trying to work through those and come to the common set is difficult.

But I think yesterday's discussion was particularly relevant. You start with the end point. You figure out what you want to achieve and then you back-map to what's the

most expeditious way to get there, and I think it depends on how much time you have, how much money you have and what your priorities are.

And so, having a really good project manager who can elicit those requirements, not only from the state level, but also, the district level and the school level, and -- broad representative panel of participants, to make sure that the system ultimately meets the need.

MS. QUELLMALZ: Yes, and I'd want to -- I would want to have some criteria that would encourage the proposers to have some kind of a survey of what's out there, so, that they are not thinking that they're starting from scratch, but they are, in fact, doing some level of due diligence, and the kinds of platforms that currently exist, you know, you don't need to reinvent the wheel, perhaps, or perhaps, you don't find the wheel you like, and so, that's your rationale.

But I would want to encourage, as part of the criteria for evaluating the

quality of the proposal, that the consortium had done some analysis of what aspects, of what kinds of architectures and platforms and inter-operability and they know what the standards are and things, so that they're starting from a basis, to make their proposal for how they're going to move ahead.

MS. WEISS: So, Randy, in the system that you described, which I thought was a very robust platform, and then if we look at some of the things that particularly Eva and Edys added, as they were talking about their's, it's really easy to see how a system like this can evolve into a full on learning management system, for a school, a district, a state.

How do you think about the issue of rich features and functions versus scope creep and know where to draw the line and what this project should do and shouldn't do? Just what advice do you have from experience, for states about how to think about limiting the scope of this in appropriate ways, as well as maximizing the impact?

MR. BENNETT: I'm not sure I have all that much good advice to offer on that. Tony might have more, having worked in the state and perhaps, experienced the temptation of a scope group.

MR. ALPERT: I've been accused of it often, at least.

Well, yes, I think it's really -- it's honing your expectations down to the minimum parties.

MR. BENNETT: The essentials.

MR. ALPERT: Yes, and the minimum essentials. So, and that's going to be driven by the criteria that are established in the RFP, I think.

What is it that you folks think is most important in achieving and then states have to wrestle then, with what are the minimum elements necessary to achieve it?

MS. BAKER: My comment would be, I would agree with that theoretically. I think practically, there are credibility elements that have to be put into a system.

For instance, hypothetically, imagine you could have a platform that operated without teacher input, just hypothetically. That wouldn't be smart for many, many reasons.

So, there needs to be certain kinds of elements that are included for -- to get buy-in from the audiences that you need buy-in for, even if they're not the most directly relevant to the main line cast sequence that you have.

So, in some ways, you can make them relevant, but I agree, we're looking at parsimony, rather than elaboration.

MR. BENNETT: Yes, well, one of the -- I'm sorry.

MS. WEISS: Go ahead.

MR. BENNETT: I guess one bit of advise I would have is, as Eva suggests, design for a parsimony, but make sure that whatever it is you design has the ability to easily interoperate with other systems that do those other things that you have not designed

your parsimony system to do, but which you suspect it ought to be able to --

MS. WEISS: Eventually expand.

MR. BENNETT: Yes, exactly, exactly.

MS. WURTZEL: So, Randy, actually, I think answered the question I was going to ask, but let me pose it a little bit differently and see if we get more explanation, which is, we're not writing on a blank canvas.

In many ways, lots of states and districts have partial, or as Joanne said, full online management systems and they're connecting assessment results to instructional material, pointers, formative assessment.

So, how should we think about -- how should states think about connecting this platform to what they have, or should they, and a way, so that you could integrate them I a fairly seamless way from the user perspective in a district and school.

MR. BENNETT: Yes, I think that's going to be a challenge and it's going to be

more of a challenge, to the extent that existing systems use proprietary formats, as opposed to common -- formats that can be more commonly shared.

So, you know, to the extent that this is true, that there are proprietary formats dominate, then I don't know enough about those systems to know whether they do or they don't. It's going to be an issue.

But one good thing you can do is begin to force movement toward those more common formats.

MS. WEISS: So, I just want to end with one question that came from the folks in Utah, who have a very concrete question that in know you can answer, Tony, and if others can as well, jump in.

But it's a difference -- the question is about the difference in cost between paper and pencil versus computer administered assessments. Their experience from doing a significant portion of their testing online over the past few years has

been that their personnel costs have been about five times what they are in the paper and pencil world, at the state level. Is that consistent with your experience? State and LEA?

MR. ALPERT: Yes, I don't know how to answer that, because our online system is so different from our paper system.

So, we have three opportunities to test, whereas with paper, we would have only had one, and so, I think there was a cost shift to some extent, from the production of paper to the use of computers and personnel.

But overall, I think there was a reduction in cost, even if the expenditures increased or stayed the same. So --

MS. WEISS: So, you can connect maybe afterwards and talk a little bit more about this.

MR. ALPERT: Yes.

MS. WEISS: I want to make sure that we have all the time that we need reserved for our public speakers. So, what we are going to

do is take a quick five minute break. Are we ready, Kevin, to shift over the podium, and stuff for folks?

So, we're going to shift the podium up. Anybody who signed up as a public speaker, should come up right here, to this part of the room and we're going to line you up in order, so that we can be as efficient as possible in this process. Make sure everybody has all of the time that we've promised them, and we'll reconvene back in our seats in five minutes. Thank you.

(Whereupon, the above-entitled matter went off the record at 11:30 a.m. and resumed at 11:35 a.m.)

MS. WEISS: Okay, we are going to go over our speakers in -- in the order in which they are CP'ed, I believe. They all their numbers up there.

We will have on your podium -- everybody gets five minutes to speak. We don't respond to the speakers. We do listen hard and take notes and anybody, the speakers,

our experts or any members of the public, I just want to reiterate, are welcome to submit written input to us that we will consider and I promise you, we're reading all of it, as part of the work that we're doing here.

So, with that, let me ask the first speaker to come up and as you're coming up, I will tell the speakers that on your podium, you'll see a timer, that's got a count down and it will start blinking -- it will start turning yellow and then red, when you're out of time. So, keep an eye on that, if you don't mind. Thanks, and start by introducing yourself, please.

MR. DOLAN: Good morning. My name is Robert Dolan. I live here in Massachusetts. I'm a neuro-scientist, a software engineer and an education technology researcher and designer.

I'm currently a senior research scientist with Pearson Education. I have two school aged children, one of whom has specific learning disabilities in reading and math.

I'd like to speak to the first question on technology and innovation and assessment, namely recommendations on how different innovative technologies should be deployed and why.

As we've already heard, the innovative use of digital technologies has tremendous potential for revolutionizing the ways in which we assess students.

I'd now like to emphasis that this potential is most relevant to the accurate assessment of a wider range of students, demonstrating broader definitions of knowledge and skills than we can currently handled with traditional testing methodologies.

New technologies provide unprecedented opportunities for flexibilities in the ways we assess. The flexibility will allow to minimize the assumptions we currently make about the most appropriate ways for students to demonstrate their knowledge and skills.

What I'm talking about goes beyond

serving students with disabilities, because for every student diagnosed with a learning disability in reading, there are many for whom reading is simply not the best way to understand a question science.

For every student who qualifies as an English language learner, there are many for whom writing and explanation of how they solve the geometry problem is just not the best means for demonstrating their knowledge and skills in geometry.

What I'm talking about goes beyond accommodations. Accommodations remain largely an after-thought in educational assessment. Yes, we need to provide students with the supports that we currently rely on accommodations to provide.

However, we need to consider this process from the outset, during initial item and test design. We must think not in terms of which accommodations can be applied during test administration to fix an item, but rather, what construct or relevant barriers

individual students might face when exposed to the item.

This includes a grounded understanding of how students interpret items, what is expected of them, how they strategize answering items and finally, how they respond to items. So, how do we do this?

I'm advocating for the application of universal design, but in a deeper way than we have to date in assessment. Universal design tells us to provide individuals with choices to minimize barriers to their successful participation in activities, barriers that would otherwise unnecessarily exclude them.

Innovative technologies provide us the means for delivering this flexibility. CAST, the Center for Applied Special Technology here in Massachusetts has led the field in applying universal design through universal design for learning, to education, by considering how digital technologies and new media can provide rich supports for

student learning and reduce barriers to the curriculum, while still maintaining high achievement standards for all students.

Specifically, a universal design approach to deploying innovative technology and assessment can help in two ways.

First, I can help determine from a student center perspective, which interfaces and which item techs are most appropriate, whether we're looking for technology to allow testing to greater depths of knowledge or to avoid the impact of construct irrelevant factors, inherent in traditional testing.

Second, universal design can help us minimize the risk of introducing new barriers to student performance, such as by subjecting them to overly complex user interfaces.

Frameworks for applying universal design to the use of innovative technology and assessment are already being developed, including for example, ones developed by CAST, the University of Maryland, SRI and Pearson.

Obviously, such a flexible student centered approach to education applies not only to assessment, but to instruction. It is here that universal design has already made the most headway in defining how to appropriately individualize curriculum, materials and instruction.

As such, it's critically important that efforts to innovate the ways in which we assess students occur in parallel with efforts to innovate the ways we teach them.

This also has large implications for learning standards. If we are to provide students with options and choices, we must empower them with an understanding of themselves as learners, and we must prepare and empower students and teachers to identify and understand what engages each student and to work together to choose wisely, the learning situations and modes of assessment that best allow them to acquire and demonstrate their knowledge and skills.

Such meta-cognitive aspects of

learning cannot be after-thoughts or lie only between the lines of our standards, but must be explicitly emphasized as core expectations.

I believe a full embrace of universal designs principles is of core importance, as we charter course for future assessment systems that rely on innovative uses of technology, if we are to truly support and celebrate the diversity of students, the diversity and ways in which they learn and the ways in which they can demonstrate their learning.

Thank you for this opportunity to speak.

MS. WEISS: Thanks so much.

MR. MINO: Good morning.

MS. WEISS: Good morning.

MR. MINO: My name is Michael Mino. I'm a Massachusetts resident and I'm the Director of the Center for 21st Skills at Education Connection in Lichfield, Connecticut, a regional education service center.

I think I'm going to continue the theme that was just started by the first speaker.

A recent study by the John D. and Catherine T. MacArthur Foundation captured a growing body of research evidence on education, technology in school aged youth, when it stated the digital media, online communications have become pervasive in the lives of youth in the United States.

Social networks, online games, video sharing sites, gadgets, such as ipods, game players and mobile phones are now fixtures of youth culture.

While students are growing up in a world surrounded by digital devices, digital media and the internet, they increasingly stifled and stymied in the same in traditional schools.

Technology and innovation are reshaping the world we live in, yet much of education persists in a mold that was cast 100 years ago.

The Race to the Top Assessment Program is an opportunity to invest in innovation and technology, to transform both the delivery of instruction, curriculum and formats of assessment.

I might ask rhetorically, how many people this morning are updating their Facebook page or doing a Twitter feed from this meeting? I found none on Twitter this morning.

While there are many technological innovations impacting our world today, I'd like to focus on just two, that has significant potential to transform education in an affordable and comprehensive manner.

Social networks and open source software are two transformative developments that are exerting significant influence on the business and entertainment sectors of our society.

Social networks have the potential to create communities of learners that truly break down classroom walls, the district

boundaries and open teaching and learning across a broader spectrum of students, teachers and expert practitioners.

The open source movement is transforming business practice and has the potential to do the same for education.

Open source development leverages expertise to produce freely distributed software that can free dollars for service and support, as opposed to licensing.

There are many openly available technologies that could be transformed for use in education, but two that I want to specifically focus on, have been developed by educators for educators.

They are the Moodle Learning Management System and the eLGG Social Network Platform.

Moodle is an open -- an online learning management system that is freely available under an open source software license. Moodle is being used successfully from elementary education to higher education

and workforce development in all subject areas across art, language arts, humanities, science, technology and mathematics.

Moodle is a robust system being used around the world. For example, the University of Brazil has 34,000 users. San Francisco State University has 34,000 users. The Austrian Federation of Ministry of Education has over 110,000 user and the UK University, open university, is well over 180,000 users.

Moodle enables students and teachers to extend learning beyond the physical boundaries of a classroom and into a 24/7 online environment that's freely available via the internet for delivery to a variety of digital devices.

Moodle has many characteristics that could classify it as a social network platform, but the educational strength of Moodle is in the ability to deliver curriculum, common instruction activities, assessments, lessons and presentations.

Moodle has a suite of collaborative tools that allow for communication, critique, discussion and evaluation between students and teachers across the environment and Moodle can be customized.

eLGG is a social network platform that's freely open, also as an open source license. It provides the functionality run of social network site that could be closed or open and it comes with advance management tools.

I'd like to give a case study example of how Moodle and eLGG have been integrated and leveraged to improve teaching and learning in Connecticut.

Connecticut career choices program is an office for workforce development funded initiative that's also funded in part by a National Science Foundation grant and State Department of Education funds.

It engages over 1,200 high school students, 50 teachers from 12 urban and 28 suburban schools in an online environment.

Moodle is developing new -- CCC is developing new courses in elective science and technology curriculum to engage and develop students skills in stem areas.

CCC builds on student inherent interest in the internet, digital media, social media and technology to engage and support them in the science and technology, in courses such as nano-technology, bio-technology, digital media and animation, virtual environments, sustainable design and computer game design.

Through the CCC program, students are grounded in science fundamentals and engage in a variety of digital media, while studying these topics in science and technology.

CCC is also articulating courses into the Connecticut Community College system and has demonstrated significant impact on student interest and ability in stem fields, and so, I see that I am out of time, and my additional notes and evidence has been

submitted I writing, so, thank you.

MS. WEISS: Great, thank you so much, Michael. Next?

MR. NIGUIDULA: Good morning. I'm number three. I'm David Niguidula. I want to focus my comments on digital portfolios.

I'm a founder and -- of a research and software development firm in Providence, Rhode Island, called Ideas Consulting.

In the early 90's, I led the first research project on digital portfolios for K to 12 schools, when I was at the Coalition with Central Schools at Brown University.

Now, I'm going to try to attempt in the next five minutes, to summarize the last two decades of work.

The questions for today asked us to discuss how we can create "better assessments". Now, assessment is a means to an end. If our end result is to improve student achievement, the "better assessment" system should result in a change in teaching and learning.

In school after school across the country, we've seen digital portfolios, when accompanied by appropriate professional development, transform the daily practice.

Okay, a quick description. Digital portfolios are multi-media collections of student work. Now, our particular system, called Richer Picture, teachers and students organize that work around a set of expectations, sort of state standards or 21st century skills.

Now, over the course of a year, a student typically enters two to four pieces from every subject area, and for each piece a student enters, the teacher can assess the work using an online Rubric, linked back to the school or state standards.

At certain points, students can create sub-sets of their work to display for different audiences, from parent/teacher conferences, to graduation committees, to potential employers.

Over a time, the portfolio provides

a powerful platform for students to use their daily school work to demonstrate standards, while showing their individual strengths in areas for improvement.

Our work has been across the country and across the grade levels. At an inner-city charter school in Los Angeles, digital portfolios store brief video samples of each students' independent reading several times a year, creating an online progress report.

Similarly, teachers score digitized rating samples against a common Rubric, providing a very different kind of data driven decision making.

Meantime, in most schools across New Hampshire, students are using portfolios to demonstrate their technology proficiency. Through the portfolio, students demonstrate that they can meet the NETS, but just as importantly, they do so by applying those skills in the context of other classes and in our home state of Rhode Island, digital

portfolios are used by high school students as part of the proficiency based graduation requirements.

Starting with the class of 2008, students across the state demonstrated their skills and knowledge across six subject areas through the use of portfolios.

Before a student could cross the stage to get their diploma, they had to have a body of evidence that showed what they knew and were able to do, and from there students can prepare reports to of their work for a variety of audiences, including colleges and career.

In our conversations with college admissions officers, for example, a tour can contain two minute reviews of student abilities, such as the ability to write beyond the one page essay or the ability to work independently, show skills related to the major.

In short, we have viable examples of schools and states implementing portfolios,

sometimes for high stakes decisions.

We're showing every day that portfolios can be implemented on a large scale. This is not a theoretical possibility. We have protocols for creating reliability and scoring performance tasks, ranging from teachers in a department to colleagues across the state.

More importantly, the portfolios are having an effect in changing schools.

In Rhode Island, for example, under the new policy, statewide graduation rates improved from 70 to 74 percent, while the drop-out rate fell. The assessment policy drove attention to graduation, and it did a couple of key things.

First, it -- portfolios expands the range of standards we can assess. Every district's curriculum asks for oral presentations. Now, through portfolios, we have a way of measuring it.

The portfolios don't have to be huge. A same piece of work can be assessed to

cover multiple areas.

Second, the portfolios create better assessments where it counts, in the classroom. Since most entries in the portfolios come from entries provided by teachers, portfolios also demonstrate the teachers work and a great deal of our professional development with schools focuses on helping teachers generate what we call portfolio worthy tasks, assignments that are linked to standards, but also showcase the type of thinking skills that are at the heart of each subject area.

It's our contention that additional portfolio can improve the quality of interim classroom assessments to the point they can provide the same data that states currently get from external assessments. Give teachers the opportunity and they will rise to the challenge.

In closing, I'm encouraging you to look a digital portfolios as a viable method for large scale assessment. Through this

work, we can move from the snapshot assessments we have now, to the richer picture of student performance throughout the year.

To borrow a phrase from Linda Darlene Hammond, who you'll hear this afternoon, we can help all of our schools achieve standards, without standardization. Thank you.

MS. WEISS: Thank you.

MS. WRYNN-HUFF: Hi, good morning. It's most afternoon. So, good afternoon, everybody. My name is Mary Wrynn-Huff and I'm the Northeast Territory Manager for Learning.com.

We're celebrating our 10th anniversary. We just celebrated our 10th anniversary at Learning.com and have only delivered web-based digital products since its inception.

We are a pioneer -- we are a premier provider of web-enhanced instruction, assessment and evaluation solutions for technology literacy, technology integration,

21st century skills and stem.

We focus at Learning.com on innovative result oriented solutions with a singular purpose, to improve student learning.

In 2006, we introduced our first online assessment for technology literacy. In addition to providing an innovative online approach to assessing student technology proficiency, we developed a powerful and comprehensive platform for the secure hosting, delivery, scoring and reporting of high stakes assessments, and just last year, we introduced our 21st century skills assessment that incorporates a project based constructed response with embedded Rubric and tools to facilitate teacher scoring.

Our success in delivering online assessments is demonstrated by our partnerships with many states, including Maryland, Texas, Arizona, South Dakota, North Dakota and Rhode Island, for statewide assessments and targeted assessment products.

So, I'm going to go through the

questions in the order that they were written.

So, the first one you're asking, how we -- what we recommend, that different -- different innovative technologies be deployed to create better assessments.

We strongly recommend that new assessment systems leverage the increasingly available technology, as all of you spoke about today, and the infrastructure in school districts.

I've been with the company for five years and prior to that, I did staff development, and this -- it's amazing, how far all of the school buildings have come in the last five years. Infrastructure, just like some of the speakers said, is not a problem.

Bandwidth, computers, mobile devices, whole-class presentation technologies and innovative assessment systems deployed online will significantly improve the quality of flexible and relevant data and reports, the ability to integrate assessment data with other school systems and curriculum resources

and most significantly, improve teacher productivity.

An online technology based assessment system also allows for test administration in a variety of settings, including classroom, computer lab, extended day and remote environments.

The second question is, you envision the need for technology platform for assessment development, administration, scoring and reporting that increases the quality and cost effectiveness of the assessments.

I will describe our recommendations for the functionality such a platform could offer.

Our experience at learning.com has shown that there are several key success factors when implementing a technology platform for assessment development. These include simple, easy to use, intuitive user interfaces for students, teachers, districts, states and other LEA's, not to mention,

administrator and parents.

Support from multiple items, including portfolio based, constructed response and performance based items, embedded scoring Rubric that facilitate teacher scoring of constructed response and portfolio based items, tools to accommodate high and low state assessment, ability to interface with student information system and grade books, ability to tag and search for items, subject, strain, grade, etcetera, items sharing, easy to use tools to create quizzes, ability to import student assessment data, student, teacher, district, administrator parent log-ons with related security, powerful and flexible reporting that allow for reports at student on up to a district or a state.

Reporting systems that accommodate pre and post test scenarios, and one that generates individualized prescriptions, based on assessment data, and that folds into question three, is how would we create this technology platform?

That already exists. Most of the components that I described just a second ago, is what we currently have available in our platform at learning.com.

Sorry, I'm running out of time, so, I'm getting off of it.

Most of it has been created. We plan to invest an additional $500,000 over the next six months, and once fully developed, we estimate our cost to maintain this, about $250,000 to $500,000 a year.

Most significantly, we plan to release this system to teachers nationwide at no cost, beginning in the summer, this summer of 2010 and the free version will include most of what I just described. Thank you very much.

MS. WEISS: Thank you.

MR. BERGER: I am Larry Berger, the CEO of Wireless Generation, a company dedicated to innovations and assessment.

I want to start by commending the Department on having the courage to invite and

inspire coherence and collective action in education generally, and for proposing with this program, to do so in assessment.

I say that this involves courage because intrenched interests and bureaucratic habits benefit from the incoherent decentralized system we have now.

Three-hundred-fifty-million is a nice carrot, but it doesn't trump all the cash and prizes in the billion-dollar state assessment market of today, and we should all be ready for the back-lash, when this program disrupts the existing market for number two pencils and the forklifts with which paper tests are loaded into scanners.

But that market needs disrupting, so, please go for it.

I've decided to focus my comments on one key issue, focusing less on what the technology can do, and why it is that the best ideas about futuristic assessments tend to stay in the lab at my company and others, while the most ordinary ones are the ones that

states end up having to procure.

So, I have four suggestions. If want a break-through, we need to procure and manage with agility. It's thrilling to see the pieces coming together for a break-through, better common standards, predictive models of student growth, research around new items types, automated scoring, etcetera, but even as these are converging, some haven't fully arrived.

There are challenges of timing. There will be unexpected failures in an ambitious program of this sort.

So, while we must move with urgency, perhaps it's time to use what software developers call an agile methodology, iteratively committing resources, scope, features, milestones to projects as we go.

Designing a policy that insists on strict compliance deadlines may show the innovations we're looking for. Instead, we may need a new kind of project management collaborative, in which everyone in the

consortium uses their real-time best judgement at the Federal, state, district level, to make decisions about what will be in each phase of release, based on what is proven to be easy and what's proven to be hard.

It's important that you are competing both a new program of assessments and a five year program of ongoing assessment R&D.

Second point is that services are essential. It's important that the procurement make clear that what needs to get created is not just assessment products, but also, services. It's the services, training, leadership, data analysis, technical assistance, evaluation that make the product succeed, make the data matter to teachers and parents, make the program one which learns over time.

Services are expensive, so, this means making it clear that the legislation expects ongoing state and LEA commitments to services and ongoing product support and

improvement, well beyond the $350 million.

The third point is to invite different strategies about openness. To build great platforms of the sort people have been talking about, with great services, might require huge investments of up front capital, perhaps beyond the amount that would be available under this funding source.

This kind of private investment requires that there is a business opportunity on the other side to earn back the investment. The winners have to win something. That can't -- they can't be required to open source all of their great ideas, especially if the Government didn't pay for all of them.

On the other hand, this is public policy and shouldn't be used to create one winner, who can lock others out or collect exorbitant tolls to get in.

We want things to be much more open, so that a great idea or a great system can be part of a state assessment, as soon as it's ready, rather than being locked out of

the existing -- by the existing testing contractor, and so, I'll jump ahead.

But I don't think there's right answer on how open and how closed, but I do think it's important to ask states to put a stake in the ground or consortia, about where they see that.

My final point is to address the procurement problem that I think we're facing here.

As an innovative company, I'd love to team up with some states and really push the limits of what's possible. I have a lot of great things in my R&D lab that I'd love to show them, but the problem is, if I feed my best stuff to state partners during the bidding process, I would in most states, disqualify myself from actually doing the work.

So, I need to stay on the sidelines, lobbying in a white paper hero there, but knowing that if I do so, my best ideas might end up in someone else's contract.

An alternative to this is that consortia of states might agree to run a competition to select a design and development partner in advance of the March competition, in order to be able to do this sort of public/private collaboration that I think is anticipated here.

I think that this has been confusing for some states under Race to the Top, in terms of choosing partners with whom to propose it's probably something you should get out in the open in this proposal.

I want to thank you for creating this opportunity and I wish you high levels of adequate daily progress, as you try to get this done by March.

MS. WEISS: Thank you.

MR. BORG: Good morning. My name is Martin Borg and I appreciate having the opportunity to speak here today.

I'm President of Measured Progress, a non-profit company, who have worked with states on their assessment programs for 26

years, and currently work on the assessment programs in over 20 states, including the NECAP Consortium and the Massachusetts MCAT program, for which we were the original contractors.

Consistent with our not-for-profit educational mission, our bottom line, and recognized by our clients, is teaching and learning.

From the start, we've worked with states on standards based assessments that are customized, inclusive, innovative, non-traditional and geared toward a variety of student populations, general, special education and English language learners.

Our states assessments have usually included a variety of testing approaches, not only multiple choice, but also extended constructive response, performance tests and portfolios, both paper and computer based.

The scoring, standard setting and analytical techniques we devised for these non-traditional formats are widely used today.

I personally have been involved with the designing and building of online assessment engines for over 10 years and prior to that, I was a district technology integration specialist and a social studies teachers. What follows a general compilation of those experiences, listening to our states, and discussions with my colleagues at Measured Progress.

I'm in an enviable position. I can basically shorthand things at this point. Dr. Bennett, I really like the way you outlined the overall parameters and recommendations for a comprehensive system. I agree that a lot of the technologies, including novel item types, automated scoring, constructive response items, have been out and are in the industry right now, both our industry and other industries.

To answer your question, is there an off-the-shelf product, the answer is, probably yes and no.

The components are likely to be

there. It's like going to Home Depot. You can pull the components of building your house off the shelves, but how you put that together, that's where all the trick lies.

I also would like to say that I really like the way you've approached this particular situation. You start with the end in mind. You're looking for the right model, and then you begin to look towards technology, but there's one thing that we've got to pay attention to, I think, somewhere in the middle there, and that is spending some time with the teachers and the students and the people who are going to end up using this system, because the best designed assessment program or platform will fall on its face if isn't very easy to implement, if teachers don't log in and recognize it as their own, and their own environment and their own things.

People are willing to use things much more, if they get more out of it, than they have to put in, and nobody ever really had to do a hard sell on the wheel, and as it

turns out, no one really went into education because they really wanted to use the very, very, very cool assessment systems that I built. They went in for other reasons, and we should honor that.

So, grounding it in teaching and learning, I think, is very important and also know that there is a whole discipline and methodology around designing products and services and platforms. That could be maybe where you want to spend some of the dollars designated for the local areas, as getting their input early on in the system, so that when it's finally delivered, they see it as their own.

I recommend that you take a few chances and a lot of other people have recommended the same things. A lot of the times I spent asking my programmers to disable functionality and online environments, so that it would be compatible with paper, and this is all well and good for certain instances.

We're looking to move towards

computers where adaptive testing and automated digital scoring and all kinds of enhanced items can be used very seamlessly, but I just want to maybe throw a few more things out there, and not that I think they're very practical at the moment, but they could be in the future.

Why not use assessment to predict the likelihood of successful teaching and learning relationships? Why not look at learning styles and teaching styles and look for combinations that might provide a greater likelihood of students having that special teacher that they remember for the rest of their lives, be able to predict the likelihood that the interchange between those two people would be very successful in the learning process.

You could also use technology to kind of go in the back one and monitor learning and teaching, to find out which methods are better for that particular student and that particular class.

I'm fast running out of time and I would like to simply say a few more things.

There's a lot of ways to solve these problems and one way I think is important is to echo some of the open source concerns.

Make sure the items -- maybe not the entire systems, but the thinking around the display level of the items is open source, so that they are easily transferred.

I really recommend and really support the idea of building these systems, so that they could integrate with student -- information systems and other systems, so as it grows and expands it's able to talk to these other systems and there are international standards that are used for that.

I view this more as a coral reef that evolves and grows over time, than a specific destination.

I thank you for your time and I'm sure we'll be able to build a wonderful

dashboard for instruction.

MS. WEISS: Thank you.

MS. SANDVIK: Director Weiss, members and guests of the Department, I'm honored to testify before you today.

My name is Todd Sandvik and I represent MetaMetrics, an educational measurement organization that works in assessments with states and districts, commercial publishers and not-for-profits.

Today, I will submit ideas on what we see as the next generation of assessment, with the goals of supporting learning and improving outcomes for all students.

First, I will suggest a unified approach to objective measurement that would categorize educational assessment and make it more supportive of the needs of educators, students and parents.

Then, I will give several examples of how this approach can promote more focus on individual learning needs through technology and innovation.

Unfortunately, today's test results rarely inform instruction, if at all. The main problem is not the reliability of assessments, rather there are too many scales in educational testing.

Instruments that measure similar constructs do not freely compare information, leading to barriers and inefficiencies in the utilization of this data by educators and parents.

To address these weaknesses, new common assessments in reading and mathematics should incorporate two key innovations in their design.

One, connect the day-to-day with the year-to-year through standards scales that are vertical and are developmental in nature, and two, report actionable information that supports differentiated instruction.

For example, a reading measure should inform reading practice across the curriculum and at home. The measurement of scales and the common assessments should be

pervasive in education. No matter the tests, results should be expressed in comparable terms when measuring the same construct.

New programs should integrate the same scales of the common assessments, while equating studies should align other assessments to enable reports in a standard metrics.

To provide actionable utility, the scale should also be used to calibrate contents, such as reading passages or mathematics lessons. Having a common frame of reference will foster differentiate instruction because students and contents -- content can be matched more efficiently.

Once the objective instruments report in a standard metric, multiple measures are no longer a burden, but a strength, as data accumulate.

Today, psychometric experts worked tirelessly to enhance the reliability of instruments under external constraints. How long can we test? How many items? Fixed form

or adaptive?

Still, no matter the quality of the test or item type, individual performances may be affected by unpredictable factors, such as whether a student simply had a bad day.

With common metrics, data systems can be leveraged to refine any estimate for a student by incorporating all available information, new and prior, year after year. Statistical methods offer such capability and can be deployed as basic computational utilities.

Not only would common metrics provide educators and parents with more and better data, but they could be used to model informative growth trajectories for individual students.

What does this look like? Take for example, a graph drawn to predict a hurricanes path. Over time, as information is added to the prediction model, when certainty is reduced and the range of the expected landfall narrows, yet, from early on, observations are

extended to likely outcomes, to inform appropriate planning and action at a given time.

A similar graph can be produced to illustrate a students trajectory toward the ultimate goal of college and career readiness. With common metrics, the standard to which all students are driving will be transparent, as would an individual student progress toward that goal.

Such learning trajectory should complement, if not superceded, today's emphasis on periodic summative outcomes. Instructors should focus on the data driven practices that most improve learning trajectories and growth, rather than arbitrary interim benchmarks.

Finally, assessment can be more closely integrated with instruction through common metrics and enabling technologies. No longer must assessment instruction be seen as competing priorities. We should pursue blended strategies to enhance student

learning.

As students spend more and more time learning online, information from computer based instructional activities can produce over time, highly efficient assessment feedback that can be used to differentiate instructional practice.

For instance, a web-based technology deployed in pilot districts today embeds computer generated reading items into on-demand periodically content.

As students engage wide ranging text, whether it's teacher directed or student directed, at home or at school, they respond unobtrusive test items that deliver real-time feedback to students, teachers and parents.

While these items in small numbers lack the reliability of conventional tests, the many hundreds, even thousands, of responses a student might produce during a single year lead to impressively low measurement error when taken all together.

The frequently updated measures

help to keep students matched with appropriately challenging text every time they read.

In summary, the initial developed common assessment should promote the unification of educational measurement. Scales should be vertical and widely implemented to accumulate vast highly useful data over time that can be tied to instructional content.

Such data will be valuable in refining student estimates at a given time and in developing learning trajectories that show progress towards college and career readiness, which should be the persistent goals of K-12 education.

Moreover, assessment instruction can be blended more efficiently to enable teaching and learning, based on individual student needs.

History points to the benefits that can be achieved in unifying an objective measurement construct, consider time. Where

would we be today if each train station still had a local time zone for calibrating clocks, as was largely the case during the late 1800's?

Similar unification in educational measurement is an imperative to give more meaning to assessment and promote widespread innovations of learning. Thank you.

MS. WEISS: Thank you.

MS. BUCHANAN: My name is William Buchanan. I'm one of the graduate students at Brown University's Urban Education Policy Program, and I first wanted to begin by thanking for the opportunity to speak here also, thank one of our research team members, Andrew Morris, who is unable to attend today.

Our research began as a class project in Dr. Kenneth K. Wong's urban politics and school governance course. It's part of the urban education policy curriculum at the Graduate School of Education at Brown University.

Rhode Island recently appointed a

new chief state school officer, Deborah Gist, who knew the state was in need of dire systemic change.

One of the five major changes that she identified has been the educational standards and assessment methodologies that are currently used within the state.

A grant from the Bill & Melinda Gates Foundation has previously funded initiative between the Rhode Island Department of Education and the educational alliance at Brown University, to develop a comprehensive high school graduation system, which you already heard some about.

However, this system was only used at the high school level, a point which makes it increasingly difficult for educational providers to intervene and correct for years of failed education.

To better prepare students, their families, and the education system as a whole, to meet the requirements of No Child Left Behind, our group investigated the use of a

two-tiered system of assessment for kindergarten through 12th grade, because beginning assessment at grade three, where students stop learning to read and start reading to learn, is also too late to affect the changes that truly need to occur.

While we agree that there are many flaws in standardized test models, ranging from test alignment to cultural relevance, to lack of output measurements, we still understand the usefulness of a broad assessment, and think that should still be implemented.

However, feeling the need to follow current trends in the social sciences, we propose creating a system that would not only capture quantitative data, in the form of standardized tests, but also, qualitative data, for portfolio assessment systems.

It may be easy to see how this may work at the state level, but there is also ample opportunity for portfolio assessment to be used as part of the common core assessment.

Developing a Rubric based on content and performance standards in the common core curriculum, the Council of Chief State School Officers can in a sense, standardize the measurement of the work.

Aside from the benefit of creating a mixed method assessment system, there are many tertiary benefits of the assessment methodology.

For example, transparency in the portfolio assessment provides more opportunity for parental and community engagement, provides administrators a way to assess how well teachers are preparing the students, to directly meet the requirements within the standards, provides students with more immediate feedback and most importantly of all, for students, it allows them the opportunity to demonstrate learning in the most appropriate and relevant method.

Based on a five point scale, student work would be assessed using a Rubric. However, Rubrics in the past have tended to

focus on positive measures for achievement and negative measures for lack thereof.

In our proposal, we not only suggest using positive measurements for each point of the Rubric system, but also suggest using developmentally appropriate indicators to develop a more robust indicator system.

By this, we mean it is necessary to have positive indicators, but there are students who may be prepared to show proficiency, but still lack the necessary developmental capacity to meet a particular score.

Using a method which includes positive and negative indicators creates a more thorough tracking system.

It is also our intention that portfolio assessment occur on a semester basis, since a once a year summative assessment isn't adequate to provide the necessary data to allow teachers to intervene.

Developing a multiple measurement system to run concurrently with it, can

provide students the opportunity to receive interventions that are needed closest to the time when they are to be beneficial.

More importantly, for the educational system, is that it truly measures student progress and provides a more descriptive and robust database of student achievement.

Along with our prepared statement, we've also provided a copy of the report that we prepared for the course, with much more detailed information about the specificity of our proposal within the document.

Again, we want to thank you for the opportunity to speak here today and hope our contribution can help the US Department of Education with supporting the common core assessment that provides equity for all students to demonstrate knowledge, develop a more transparent system, which parents can feel more engaged with and a system that allows teachers to be held accountable for student outcomes that directly measure student

performance and knowledge.

MS. WEISS: Thank you so much.

MR. HOFFMAN: My name is Tom Hoffman. I'm also from Providence, Rhode Island and I'm a technology consultant specializing in student information and assessment systems. I'm project manager of School Tool, which is an open source administrative platform for schools. I also work with the CANDU project, which is an open source competency tracking application, used by career and technical centers in Virginia, and I am former English teacher in the Providence public schools with a Master's in teaching English from Brown University.

I'd like to recommend some specific facets of the technology platform for assessment, particularly in reference to Race to the Top criteria, addressing accessing and using state data to support decision making and continuous improvement of efforts in such areas as policy, instructions, operations, etcetera and also, making the data from

instructional systems together with statewide longitudinal data system data, available and accessible to researchers.

These requirements suggest a high degree of data portability, interoperability and integration with aspirations for complex data warehousing, business intelligence and inferencing expert systems.

One of the technical foundations of this type of platform is the development of ontology, which we heard about this morning from Dr. Baker, and to review, can be defined as a formal representation of set of concepts within a domain and the relationships between those concepts.

The potential role of ontology as an educational research and throughout the implementation of educational technologies and data systems parallels their growing role in bio-medical research.

I would specifically propose funding the creation of a national center for educational ontology modeled on the national

center for biomedical ontology, which is funded by the National Institute of Health.

From their website, the goal of that center is to support biomedical researchers in their knowledge intensive work, by providing online tools and web portal, enabling them to access, review and integrate desperate ontological resources in all aspects of biomedical investigation and clinical practice.

The center is funded by the NIH roadmap for biomedical researchers, bioinformatics and computational biology initiative.

The roadmap was launched in September 2004, to address road blocks in research and to transform the way biomedical research is conducted by overcoming specific hurdles or filling defined knowledge gaps.

These are programs which might not be otherwise supported by the NIH IC's, because they're scope or because of their scope and because they are inherently risky.

With the consistent ongoing commitment to the development and use of ontologies, the National Institute of Healths' recovery act fund has already supported 61 current research projects using or contributing to biomedical ontology.

By comparison in education, despite contributions from a desperate set of actors, including CRESST at UCLA and a 501(c)3 education research organization, there is no central hub for research and development of or use of ontology. Individual projects tend to emerge and disappear and there is in particular, no commitment to the kind of open and collaborative environment that now simplifies biomedical ontology.

For example, the National Forum on Educational Statistics at the Department of Education has created a national educational data model. It is similar to an ontology, but the data model is more constrained and potentially much less rich and powerful than an ontology approach.

However, it would be an obvious foundation for the development of a subsequent set of educational ontologies.

CRESST has developed several detailed domain ontologies, as you saw, for specific subjects, such as algebra, as part of their research. However, unlike their peers in biomedical research, publishing, collaborating and promoting these ontologies does not appear to be a priority, which limits their influence and impact.

Similarly, I can see from their presentations that CRESST has developed a tool called CRESST knowledge mapper that looks quite useful, but does not seem to be publically available, either commercially or for free, and thus, does not contribute to or promote further development of domain ontologies in education.

In contrast, the National Center for Biomedical Ontology protégé editor is an active and prosperous open source software project that has become an industry standard

application.

As was the case in the biomedical field, an investment in educational ontology is relatively high risk and does not fit, obviously, into existing programs. However, if we don't start this process now, while we have this unique wind fall, I don't know when we will.

Rest assured, however, that this essential, foundational research, given the vast ambition for educational data system, ontologies will become integral to educational research and applications used in schools, as they become in the biomedical field, and sooner or later, the value of our solutions will be bottle-necked by the quality of our ontologies.

MS. WEISS: Thank you.

MR. STAHL: Hi, my name is Skip Stahl and I'm batting clean-up this afternoon. So, but I realize I'm competing with lunch time. So, I'm going to make this brief.

I'm currently senior policy analyst

and one of the founders of CAST, here in Wakefield, Massachusetts, and presently, I'm project director for the National Instruction Materials Accessibility Standard, NIMIS Center and Chair of the NIMIS Standards Board. So, I'm a big component of standardization, particularly when it comes to educational content and materials.

Assessment is one component of instruction and indelibly linked to the three other -- to three others, goals, methods and materials.

We propose the development of a digital media assessment based on the framework of universal design for learning, where students must be provided with one multiple means of recognition of assessment, directions and stimuli, two, multiple means of interaction and expression within assessment tasks and three, multiple means of engagement during the assessment process.

With proper design, implementation and validation, these principles can guide the

development of the next generation of summative assessment evaluation instruments, resulting in a more authentic and accurate measure of the achievement of all students.

This design from the outset approach will also eliminate one, costly retro-fits, two, ad hoc accommodations that may violate test validity, three, a large portion of alternative assessments, and in particular, I'm thinking about the two percent modified alternative assessment, and four, other interventions that require student failure before interventions can occur.

So, what are some of the current challenges? In some assessments, reading is the target scale being addressed. In other assessments, reading is required, but the area being assessed may be mathematics, history, science. In assessment lingo, reading and these assignments is construct irrelevant.

For students in the margins, those with disabilities, English language learners, students from different cultural backgrounds

and others, these construct irrelevant demands generate artificially low achievement scores.

Confounding vocabulary introduced in a math assessment may pose a significant barrier for students in the margins, undermining their test performance, regardless of their proficiency, with the subject or skill area being tested.

What have we done to address some of these issues? Accommodations are the preferred means by which the media and/or methodology of assessments are retro-fitted. Realistically, the extent of accommodations required to make a test accessible to a student with a disability is a direct reflection of the inappropriateness of the instrument for that population.

Accommodations should not be touted as solutions. They are indicators of narrow and non-inclusive assessment design.

The solution is not alternative assessments. The solution is assessment sufficiently flexible and valid, designed with

all students in mind.

Another challenge that comes up, particularly with students with disabilities, is the need for continuity. IDEA 2004 notes that assessment accommodations should be chosen on the basis of individual student need and generally be consistent with the accommodations provided during instruction. This rarely occurs.

Computer based text-to-speech offers independent text access for students with print disabilities, and is commonly used during instruction, but rarely available during large scale testing.

So, what are possible solutions? True solution, design assessment systems differently from the start, creating them from the outset to be accurate for the widest range of students. In order to promote this, we've developed two sets of guidelines.

One, universal design for learning guidelines, available to national center for universal design for learning, and the other,

in collaboration with Pearson Educational measurement, even though the guidelines themselves are vendor agnostic, there's a new set of guidelines published in June 2009, called universal design for computer based testing guidelines and we hope that they will prove to be resources for you and for the Department, as it makes its deliberation, moving forward.

A universal design for learning approach can offer guidance for enhancing student engagement and persistent flexibility in recruiting attention, sustaining effort and supporting self-regulation are all highly individualized, nearly impossible to address without employing the magic of digital media.

However, the components of adaptive testing often point to the automatic difficulty adjustments if that approach is enhancing student engagement by decreasing the challenge presented to them. This is the same rationale used to support the simplification of the curriculum for struggling students

identical often to out of level testing, that resulting in moving students with disabilities further away from the mainstream curriculum.

Universal design for learning seeks to maintain high achievement standards for all students through the use of customized scaffolds and supports, that re-enforce the importance and maintain grade level expectations for all learners. Thank you.

MS. WEISS: Thank you. So, now, I think I'm what's standing between you and lunch. So, I will wrap it up quickly by thanking all the people who came to us as public speakers today, to share their thoughts. We really appreciate it. To all of you in the audience, who took your time this morning, I hope it was a valuable use of your time, and you learned something that you didn't know yesterday, and to Tony, Eva, Randy and Edys, I want to just sincerely thank you on behalf of the Department, for sharing your time and your wisdom with us. We truly, truly appreciate it and I'm hopeful that it will

make the notice better than it would otherwise have been. So, thank you so much for coming.

For those of you who are joining us this afternoon, for the high school panel, it will reconvene in here at 1:30 p.m. Thank you.

(Whereupon, the above-entitled matter went off the record at 12:30 p.m. and resumed at 1:30 p.m.)

1