THE BROOKINGS INSTITUTION

STANDARDIZED TESTING AND THE COMMON CORE

Washington, D.C.
Thursday, November 29, 2012

PARTICIPANTS:

Introduction and Moderator:

RUSS WHITEHURST
Senior Fellow
The Brookings Institution

Keynote Address:

DAVID COLEMAN
President
College Board

Panelists:

MATTHEW M. CHINGOS
Fellow
The Brookings Institution

JEFFREY NELLHAUS
Director
PARCC Assessment

GERARD ROBINSON
Former Commissioner
Florida Department of Education

NANCY GRASMICK
Former State Superintendent, Maryland Department of Education
Presidential Scholar, Towson University

* * * * *

P R O C E E D I N G S

MR. WHITEHURST: Well, good morning, everyone. Welcome to the Brown Center Brookings event on the cost of assessment. Unless you just walked in off the street because we have good pastries and free coffee, you probably know that the nation is in the middle of a transition from an accountability and testing regimen that has been state based to one that is perhaps still state based but represents cooperation among states to create a new set of common standards and aligned assessments. What we're going to be doing today is talking about that transition, focusing first on the information we've developed here at the Brown Center in terms of what it cost to develop and procure assessments, and then addressing more widely the quarter of the Common Core, what's coming, what the challenges are, and what's next.

I'm excited by the event because of the panelists and the speakers that we have. We're joined today by David Coleman who is the new president of the College Board and was previously a principal player in the development of the Common Core State Standards, certainly the English Arts portion of those standards. Then we have three former state chiefs, education commissioners, Nancy Grasmick who is now presidential scholar at Towson University, and she was a former state superintendent in Maryland for 20 years; Jeff Nellhaus who was directing one of the consortia that's producing the new assessments, but was formally deputy commissioner in Massachusetts; and Gerard Robinson who's going for the record of being commissioner almost everywhere, so he's been commissioner in Virginia and Florida.

So the drill is we're going to hear first from my colleague Matt Chingos who has done some of the research we're going to be talking about today, and then Matt will introduce David Coleman to give an address. And then the panelists and Matt and I

will come to the stage. We'll have a little conversation, and then we will throw it open to questions from you. I'm asked to remind you to adjust your devices so that they do not make noise, but don't turn them off, because we'd love for you to live Tweet what we're doing, and that's the Tweet address on the board.

So let me introduce Matt Chingos who's a fellow here in the Brown Center. He's our go to guy on a variety of higher education issues, and he's been doing work on the cost of assessments. It's been generously funded the Lumina Foundation. And we're all interested in what Matt has to say. So: Matt Chingos.

MR. CHINGOS: Thank you, Russ. Good morning. So before jumping into a brief overview of the report that we released this morning, I'd like to thank my colleagues who assisted me on this report. Mike Gallaher, Christine Lai, Emily Russel, Diana Stockwell helped with the enormous data collection effort that underlies the report. They went through literally thousands of pages of contracts between assessments vendors and states that we received on DVDs and boxes, all sorts of forms, to extract the data that went into the report that I'm going to be talking about this morning.

So we took on this project because of the big changes that have happened to standardized testing in the past decade and the changes that are now on the horizon. So states have expanded their assessments systems since the passage of No Child Left Behind in 2001 required them, in most states, to expand those systems to cover more students in more grades, in some cases in more subjects. And it's not surprising that these expanded systems have cost more, and those costs have come under increased scrutiny for at least two reasons. First, the budget pressures that most states are facing now have caused them to question the cost effectiveness of everything they spend on. And second, testing and the way that test results are used is itself

controversial. It's so controversial that we had to set up a medical detector outside the event today.

So the coming years will see the biggest changes to state assessment systems since No Child Left Behind, because states are going to need new tests as they adopt the new Common Core State Standards. We don't know all that much about how much the current assessment systems cost, much less about how much these new systems are going to cost. So for our project, we will were successful in obtaining data from the contracts between states and their testing vendors for the primary assessments, math and reading, in 44 states and the District of Columbia, which account for a combined $669 million in spending on these primary assessments systems. We focused on these contracts, because although they're not the only source of spending, at the state level they, for the most part, make up the lion's share of spending, and they were data we could collect in a way that's comparable across states.

So the $669 million across these 45 states amounts to $27 per student in grades 3 to 9. Those are typically the grades tested under NCLB, three to eight, and then one grade in high school. But that number varies quite a bit by state. On the lower end of the range, you have states like Oregon, Georgia, and California that spend in the range of $13 to $16 per student on these contract costs. And on the upper end some higher spending states, you have states like Massachusetts, Delaware, and Hawaii, $64 to about $105. So this total is contract spending, primary assessments, 45 states. We also make a rough back of the envelope type calculation for what if we estimated what the six states that we couldn't data from, if we included them, if we estimate spending on tests other than the primary ones, tests like science, reading, high school tests, English language learner tests. And then also non-contract costs, most importantly, the important

work that's done by state officials in state assessment offices in selecting contractors, overseeing the implementation of the tests and what not. So based on that calculation, we estimate about $1.7 billion per year. So that seems, I think, like a lot of money, but in the context of a $1,658 billion public education system, it's more or less a drop in the bucket, about one quarter of one percent of spending each year.

So if you want to think about what that means, what you could do if we were to stop assessment activities entirely and take that money from the state level and roll it into other uses, we could decrease the pupil-teacher ratio by point one students, or we could increase every teacher's salary by one percent, about $550. So this relatively low level --

MR. WHITEHURST: Don't offer the second part in that.

MR. CHINGOS: Fair point. But so this low level of spending coupled with concerns that the quality of tests in many states is not high enough, especially if they're going to be used for high stakes purposes like holding students, schools accountable, using them as part of teacher evaluation systems. So that quality isn't high enough actually suggests that states may be under investing in their investment systems. But at the same time, more funds are unlikely to be forthcoming as states are under pressure to reduce spending. It's not an easy time to find money to do new things, even if you think they're important. So states need to find efficiencies in order to absorb budget cuts without compromising test quality, or to free up resources that could be reinvested in upgrades to assessment systems.

So a key finding of our report is that states can save by collaborating on common assessments so that the fixed cost of test development are spread over larger numbers of students. And this is driven by this finding that larger states may also have

more negotiating power with contractors, so it's not surprising that larger states spend less per student than smaller states. But what our data allow us to do is actually put some hard numbers on what the predicted savings would be to forming a testing consortia. So for example, our data predict that by joining a 1 million student consortium, a state with 100,000 students saves 37 percent of its annual contracted testing costs, or about $1.4 million. A larger state, about 500,000 students, would save 25% or $3.9 million. You know, of course, most states have already done this for their math and reading tests. They've joined the Common Core consortia, PARCC and Smarter Balanced.

So our results certainly suggest that this effort should save money on average or allow for an increase in test quality without a commensurate increase in cost. But an important caveat is that our model can't predict the exact to the Common Core consortia, because first they're doing a bunch of things that a lot of states are not currently doing, such as computer-based assessments. And second, they're huge. They contain states with about 11 to 13 million students in grades 3 to 9, whereas we're focused on states up to $1-1/2 million, because that's where most of the states in the country are. So there certainly should be cost savings. We just can't really predict what they're going to be above and beyond what we already see in some of the larger states in the country.

It's also not clear whether it's better to have two large consortia or a larger number of smaller consortia that can specialize more. And one recommendation of the report is that states should consider forming these smaller consortia for their other tests, so as they proceed with the Common Core consortia, they could conduct some parallel experiments maybe on their science tests on some other high school tests to see

sort of the strengths and advantages of these different kinds of models.

A final conclusion of our report is that states need access to better information on assessment pricing. We saw in the contracts we collected just a wide variety in the amount of detail that's provided. In some cases, it's pages and pages of line item information, and in some cases it's just a couple pages of you're getting a testing system and it's going to cost you $10 million. And this is going to be especially important as states make their final decisions about transitioning to the Common Core assessments, because the first operational tests are scheduled for at least six months after federal support for the development ends. So they're going to need to have good information on pricing in order to figure out how to make that transition and what they want to do. The problem, as I mentioned, is that assessment pricing is very opaque. So we recommend that larger states and consortia of states use the market power they have by representing such a significant part of the market to encourage test makers to divulge more details of their pricing models. This will enable them to make more informed decisions about their assessment systems as they seek to balance the need for high quality assessments with other demands on their budgets.

Just a quick overview of the report. I encourage you to pick up a copy outside or look it up on the web for more details. It's now my pleasure to introduce our keynote speaker, David Coleman. He's the perfect speaker for our event, because he's an authority on both of the key topics today, standardized testing, as the new College Board president, and the Common Core, being one of the key players in the development of that effort. So please join me in welcoming David Coleman.

MR. COLEMAN: Hi. It's a rather busy time in my life, I must confess. I started as president of the College Board on October 15th, and going from an

organization of approximately 22 people to one of 1,400 has been a little bit jarring. But there are reasons -- excuse me?

MR. WHITEHURST: To them or to you?

MR. COLEMAN: Exactly, exactly. The reasons I'm here is I have a kind of deep respect, I think, more than some people know for some of the people that are here. I have a surprising -- it's hard to describe it any other way than -- a kind of research crush on Matt. Why? Why is this? I'm recently married, so Matt has no fear that this will become inappropriate, but Matt is the co-author with the great Bill Bowen of a book called Crossing the Finish Line. And there's a chapter in there about a phenomenon called undermatching in which students who are highly academically ready and able, shockingly low income kids, make choices that are not to their best well-being. In North Carolina where they looked at this work, when even the economics was similar across institutions low income students didn't even bother applying to the range of institutions that they could apply for. I'm not trying to say that it's always better, let's say, for a low income kid to go to the more selective institution they can, but it is a massive social injustice in this country that 70 to 80 percent of low income kids who could go to highly selective institutions don't even apply. There's nothing right about that. His research has awoken me. I am no Kant, but there's the great moment where Immanuel Kant says he awoke me from my dogmatic slumber. That was Hume. And it really is amazing when a number, when a fact, all of us who have been working in this field for some time, it is exciting when a researchers shows you something and it arrests you. And you say how can this be?

It turns out that Matt's idea that there are people who can go but do not go echoes throughout our education system. It's true of AP taking, so we have data at

the College Board that disproportionally minority kids and low income kids who are ready for AP don't wind up taking it. So if you have ten kids who are ready for AP in math and ten of them are equally ready based on PSAT, six Caucasian students will wind up taking the math AP course, four Asian students, three Latino students, and two black students. They're equally ready. They're not getting the opportunity for that rigorous work that will advance them. You see it before kids choose colleges. We see on our SAT data 40,000 students who score very well on SAT which means they probably have good grades, too. They don't go to college altogether, much less not choosing a college, right. This was found separately by the wonderful strategic data fellows at Harvard where they saw this affect again. I tell you because I really mean this as a tribute to Matt. He's the kind of guy who sees things that then opens avenues of action. I think these are facts that can and must change, and it will be a central part of my present seat at the College Board to say we at the College Board cannot stand by and watch this occur. We have the data. It is our mission that kids move on, that those who can go will go, and we're going to do something about this. And I want to thank you for helping to bring it to my attention and make it happen.

And I see in this report, like this is so unglamorous, like let us pause, really, to consider how hateful this task was. Have any of you read a single assessment contract in your life? I mean, I am like the nerdiest guy in the world. I've read more state standards and that was grim work, my friends. But assessment contracts, really?

MR. WHITEHURST: Horrible.

MR. COLEMAN: Horrible, right. I mean horrible, right. So who is this? Like what kind of self-mutilating attempt is this? But let's think about what he's done with some of these numbers, because it's so interesting when he shows that with the debate

about testing in this country. It costs too much money. Absolutely wrong, right? It is dangerously trivial in our -- let's just call a spade a spade. It is an outrage that we are basing this much on assessment and not investing in high quality assessment. It's stupid. It's wrong for teachers. It's wrong for kids. It's an outrage. But that outrage is fueled by misinformation, right? None of you would have known it was point one -- who knew, if I may, that it would change teacher student ratios by point one percent before he announced that information. Who could have done that math on their own? That's insight, right? That's real deal. That's numbers showing us something we could not have said before.

With the whole political community around the Common Core, there's a debate raging in this society about how much will it cost the Common Core to implement, right. That's one of those interesting things. But who's dared show, look folks, it may be cheaper, like sometimes doing things together is actually more efficient than doing them each on our own. I haven't heard a single, right -- the only debate in the public sphere has been how much. We in education, like, we've got to get a grip, folks. We are so bad at cutting back on what we've been doing to invest in smarter stuff. If we cannot build consensus around making smart cuts in unproductive investments, we're done. So I would just compliment his sense that there's likely not new money to building a political sense and a courageous sense that things that do not work and that we must invest in that have demonstrated the ability not to work, we've got to disinvest. Does that make sense? And invest in stuff that matters.

I say that because after so much praise for one researcher, I thought it only fair to talk about a researcher I don't like, which is Russ Whitehurst, of course.

MR. WHITEHURST: Don't believe that.

MR. COLEMAN: Yes, please don't, because I'm going to go in a different direction. You know, it's interesting. I, in the position of developing the Common Core, heard a lot of feedback from various sources, and I will admit that when someone of Russ's caliber wrote his skeptical piece about concern that we do not over promise about the impact of the common standards on their own, it affected me in a pretty profound way. Why? Because I think Russ has been a dramatic and serious force in education reform in this country. His leadership, and again I observe from afar, you wouldn't know that I know all this, he really drove the idea that random experiments and proper experimentation. He has transformed the federal government's role for years to come, forever, in the kind of research that is seen as acceptable. He turned it from rookie stuff to serious stuff. I so admire you for that. And so when he says, I don't see the data that is a serious concern. So I just want to say honor that, and I only have one beef with him. A man so might should have built the "what doesn't work" clearinghouse. That is, we must use research not only to say what demonstrably works, but we've got to start using research to say we've been investing this much money in this for an awfully long time and there is zero impact for children. So I only fault you for that lack.

But I also just want to say for a moment that the other people on the panel, you know, Nancy who drove such depth and excitement and variety in assessment with the MSPAP assessment. She's been through all the generations of this. But the depths she achieved by deepening assessment is, in my judgment, part of why she got the NAEP results. She invited teachers through her style and her long leadership into a depth of instruction that is rare in this country.

I've learned so much from Jeff Nellhaus about assessment design. I've spent hours over the years with Jeff. He was like a little rock star to me, so I would go

bug him as a kid and waste his time. And so this is one of those situations where a consultant who's read your watch then tells you the time. So the notion of me standing in front of Jeff talking about assessment is rather ridiculous, but he'll at least have the pleasure of seeing many of his own ideas reflected back at him.

And Gerard is notable, not merely for being in two states, but for the courage he's brought to all the positions he's shared. And courage, unfortunately, in our system is not always tied to longevity. Would that it were so.

And finally, and then I'll jump into it, I do want to say that a mentor of mine is here, who I think has done more to drive thoughtful teacher engagement, the hardest work of all, which like teachers basically think typically of assessment that it sucks. And so if you are a leader of teachers, assessment is typically not your friend. You know what I mean? It's easy to rail against assessment if you are a leader of teachers. If you are a leader of a union of teachers, it is particularly so. But David Sherman, who's in the audience who's a longtime mentor of mine, has really with great principal led the dimension of the American Federation, the finest parts of the American Federation of Teachers, which were competed to deeper learning to common high quality standards for kids, for knowledge in kids learning, for thoughtful assessment driving, thoughtful action, and I really applaud him for that. I'm so glad he's here. That's the hardest work of all.

Now I want to start where I just ended which is we must face reality in this country, which is that teachers have radically lost faith in assessment. That, in my judgment, is a substantial crisis, because if teachers lose faith in assessment, parents shall lose faith in assessment. People trust their teachers. We have a need in this country to redeem assessment in the hearts and minds of teachers and parents,

otherwise the accountability systems we are building will never have the depth of support they need. The differences between teachers are real and must be acted upon, but in order to build the consensus to do that and do it in a humane and principled way, we must have measures that teachers can own as a proper view of their students growth. If teachers don't believe that an assessment measures their kids' growth, how could they ever believe it measures their contribution? We must understand that the doubt, that the skepticism is a radical one. It is that these assessments don't tell the truth about which kids are getting better. Do you see what I'm saying? They don't even believe the first; much less can it judge my own input. It doesn't judge my kids' growth. That must change. It is interesting. So as I think about our agenda, it is not merely to invest in a -- what we must achieve in this next generation is not to continue our mere efficiencies, not that you're saying that it's so, it is this is our moment. If we do not rescue assessment, then the whole consensus will fall, and it won't change behavior, because if people don't believe a measure is a measure of their true work, they will resist in every quiet and obvious way.

Interestingly, in my new position, the situation -- so basically when you ask teachers what they think of testing, it's either 20 percent or 25 percent support, right. You know, that's a typical number, depending on which survey you look at. However, it flips with AP. About 80 percent of teachers think AP exams make sense. Interesting, right? And are a fair judge of their students' growth and their work. Fascinating. So I will tell you I'm hell bent on trying to think about what principals of those assessments we can bring to help work with the consortia, to build exams that teachers can trust as fine high quality expressions of their kids' work and proper examinations of their growth. And I do that not to avoid proper teacher evaluation and accountability, but rather to make it

possible, a humane system in which teachers can be a part.

And I'll say another political and social thing that I want to be very clear on. I, unlike other similar institutions, as president of the College Board, have said publicly that I shall not compete with the assessment consortia. That is, I shall not go to states and say this effort will fail by my test, period. That means it is incumbent on me as president of the College Board to do anything in my power directly or indirectly to help this effort on our country's behalf succeed, period. My job is not to protect SAT or protect any given instrument. It is that kids must do college ready work on their path and career ready work so many, many more of them are truly ready for the demands of college and career. That is my commitment and that is our institutional commitment. When they chose me, they made a terrible error, and that was part of the consequence of their decision.

I do not mean then to retreat from making our instruments excellence and worth of your use, but they must be worthy of you. They must shine and be worthy of imitation, so what I want to say is if we want to redeem assessment, I want to just set out some principles of what I think it would be to make it valuable fair. And these are principles, again, that I've been talking Jeff at for years, and I think just as he's representing PARCC here. I think PARCC is doing some beautiful work to design assessments that seriously recognize two ideas. One is that assessment is an extremely powerful signal for instruction, but you've to own it. You've got to cut the [expletive] when you're like, ooh we wrote this test and all these people are doing test preparation. They shouldn't test preparation. They should look at the standards. I mean, is it a -- like [expletive] you, like no. I hate that disingenuousness. If you put something on an assessment, in my view, you are ethically obligated to take responsibility that kids will

practice it 100 times. So when I look over an instrument like SAT, I want to say to myself, is it worth it. Is this work worth doing? There should be evidence that this work worth doing will truly prepare you for the work on the best we've got. It's fragile evidence, but we should use the best we've got. There should be evidence how people respond to it, that is, if you think it will inspire good practice, but turns out it doesn't, you've got to rethink. But PARCC has committed itself in a way that I find quite exciting, to design an assessment that is worthy of imitation, that as much as possible is worthy to imitate and take responsibility for it to imitate it, to make an analogy for SAT.

Right now, I think there's a breakthrough that the SAT added writing, because we do want to make the claim that kids need to write to be ready. Like, duh, right. To be ready for college and career, it obviously includes writing. But I have a problem with the SAT writing. So if you look at the way the SAT assessment is designed, when you write an essay even if it's an opinion piece, there's no source information given to you. So in other words, you write like what you're opinion is on a subject, but there's no fact on the table. So a friend of mine tutors in Hong Kong, and she was asked by here Hong Kong students, where do you get the examples for the essay? She said, you know, it's the American way, you make them up. Now I'm all for creativity and innovation, but I don't think that's quite the creativity we want to inspire in a generation of youth. That is, if writing is to be ready for the demands of career and college, it must be precise, it must be accurate, it must draw upon evidence. Now I think that is warranted by tons of information we see from surveys of college professors, from evidence we have from other sources, so I think there is good reason to think about a design of SAT where rather than kids just writing an essay, there's source material that they're analyzing.

I think when you think about vocabulary on exams, you know, how SAT

words are famous as the words you will never use again? You know, you study them in high school and you're like, gosh, I've never seen this before, and I probably never shall. Why wouldn't it be the opposite? Why wouldn't you have a body of language on the SAT that's the words you most need to know and be ready to use again and again? Words like transform, deliberate, hypothesis, right? Oh hi, Chuck, I didn't see you. That academic language is the real language of power, that if kids master it, they will be ready in whatever field they choose enter, particularly valuable by the way, for our English language learner community to gain that access to language, but valuable for all students. So if you take responsibility in that sense for what you're examining. You see what I'm trying to do. The math in the SAT, if we have evidence that some math matters more than other math in preparing you for the demands of career and college, we should then in the SAT or other similar instruments radically emphasize that math.

So I'm going to step back from this direct discussion of assessment to talk a little bit about the common test in the language of the common standards, but I want to also say one last thing about it. Remember the beauty of assessment that we must all remember is that assessment is not just a gate. Assessment is a way to reveal someone's potential. The beautiful idea of SAT at its most beautiful, at its most original, was that merit that was hidden behind societal walls would be revealed by a test, not that it would be a wall, but it would be a light. When we talk about the PSAT delivering kids who have AP potential, right, wasn't that interesting, that data? Many people would think in this society, oh, we're trying to get all the black kids into math, so it must be all the black kids who are ready for AP math get into AP math, because we've made that a priority as a society. Oops. Data tells us otherwise. Did you hear that data earlier? With all this supposed work we're doing to prepare African-Americans to fully participate in

engineering and leadership and math and science, two out of ten who are ready actually take AP. Oops. Sometimes assessments are a lot more fair than humans are in judging talent and seeing it. When we talk about the unfairness of testing, my friends, do not ignore the cruelties that exist between humans. Shifting back to a world where we trust our instincts is often very unkind to certain people. In fact, when you look at the statistics, it is inevitably unkind to certain groups of people over and over again.

So I want to be serious about thinking about designing assessment such that assessment is not a final step. Like, who cares if you take PSAT in a sense? But does it propel you somewhere? Does it propel you forward? Does it propel you in the right kind of remediation if you show up? So as president of the College Board, I'm extremely interested in the ethical result of an assessment. Am I making sense to you? What happens next? What is the result here? What does it drive? What happens after it?

Finally, about the Common Core. I think a previous thing that we have to think about as we develop the next generation of assessment is it is not okay if these exams are built on crappy standards. Right? Because there's a problem here that, like yeah, it's efficient to have states working together on assessment, that's nice. But if the underlying standards are not work worth doing, this could be a step backward rather than a step forward. So I want to talk about with full acknowledgement of the skepticism, I want to share with you a couple of reasons why I think these standards could be work worth doing, have potential. And one reason I took this job is I actually seriously appreciate Russ's claim that without transformations, productive transformations of assessment and curriculum, the notion that standards alone will save us is foolishness. So I change my life. I took a job where I could have some role trying to make curriculum

and assessment better, rather than ignore the evidence before me. So let's talk about whether these standards, not can they improve achievement on their own, which is foolish, but do they have the potential to advance these goals of assessment worth doing.

In the literacy standards, just to get quite specific, there are three or four basic shifts I want to share with you that underlie the assessments being developed. The first one is building knowledge through content rich non-fiction. So in elementary schools, overwhelmingly, what has been assessed on exams and in basal curricula have been stories, often very content like, to be frank. If you read, through a typical basal, it is probably not as bad as state standards or assessment contracts, but not much better. Whereas, overwhelmingly, research has shown is that the coherent knowledge kids build in early grades and the vocabulary attached to that, is a critical predictor of all their academic success, but particularly their ability to read more complex text with any depth going forward. So what the common standards do for the first time is they demand in elementary school an equal balance. So rather than kind of an 80/20 stories, an equal balance of reading that is about literature and story and myth and the proper knowledge of literature, but also about history, science, and the arts. And the further demand is that knowledge is built in those disciplines coherently, that it builds on one another, right. So that, again, the common standards clearly are not standards for history and science, but as standards for literacy, they require a shift. In the diet of reading in kids in elementary school, that's a major one. Assessments that reflect that are going to look different, right. You're going to see elementary assessment that shows kids' ability to read across a range of text.

In middle school and high school, what that shift means, as all of you know, is these are standards not just for English language arts, but also for literacy in

technical subjects, science, history, and social studies. They equally demand that teachers in those disciplines build knowledge by reading high quality source materials, whether that's primary and secondary documents in history or scientific texts combined with experimental results, bringing different kinds of data together, almost something like the analytical report we see before us which combines data and text to make arguments. That might be a social studies or scientific endeavor. So you're going to see an exam that's faithful to those standards is no longer going to be just an English language arts exam in 6th through 12th grade. It's going to be a literacy exam, an academic literacy exam where you demonstrate your ability to read high quality fiction and literary non-fiction, but also to examine scientific source materials, as well as history and social studies primary and secondary documents and analyze them. Why? Because overwhelmingly, it's clear that demonstrates readiness for college and career. Right? So that's a shift. Do you see what I'm saying? I'm talking about the most basic level, at the highest level.

The second shift is an emphasis on evidence in reading and writing about literature, as well as non-fiction. So it might be surprising that reading and writing are based on anything else, but as we've been discussing, the emphasis on personal narrative alone at the expense of all other writing is a big fact of our schools. Who do we blame, right? How could it be that we've built a K through 12 system that seems to believe if you can write a compelling essay about your personal life, it's going to advance you in college and career? Who told them that? Who told them that? College admissions did. What essay is the high stakes essay in your life? Just a personal essay. Who told them that? Even the great NAEP test, its essays have not tethered in factual information. What's your favorite day? Who are your heroes? Blah, blah, blah. Right?

When do you ever again get asked those questions in a high stakes way? Now the common standards do develop narrative art throughout the grades, including creative writing. I want to be very clear about that, and it's a lovely and powerful thing. And as you grow older, you use those narrative arts also in the service of making compelling arguments and conveying information. But if you cannot write an argument with evidence in preparation for college and career and convey information clearly, you are sunk. Hence, assessment can and must change, and I'm in a discussion with the admissions officers of this country as the president of the College Board that the sole reliance on the personal essay to get into college is a real danger. Can we complement it with an analytical essay that demonstrates kids' ability to command evidence?

Finally, as you know at the core of the standards in literacy is the notion that the complexity of the text that kids are reading is a very powerful predictor of where they'll be in college. And so what's new in these core standards is we set annual demands for the level of text students are reading. This is perhaps one of the issues on me and my great partner in this soup, Penn and Teller, are most passionate about. These standards say that all kids deserve and must read the good stuff. We must stop watering down text. We must give demanding text at every level. That's true for English. I was in a wonderful panel of English language learner leaders in this country who demanded that these students have a right to rigor. I was with the lead publisher of English language learner materials in history and science, and he said to me we publish for these kids picture books in history, mostly. Are you saying that's going to have to change? And I was like, yeah. That's going to have to change. So these core standards make a demand that all students and with adequate practice master greater rigor, and so again, in assessment terms you're going to see growing demands of rigor in assessment.

In math, the core shifts, and I'll begin to speed up here. Our rigor, excuse me, focus, coherence, and rigor. In my personal judgment the great gift of these standards based on the research of Bill Schmidt that again I know there are skeptics of, but I am personally convinced that we have enough data that there's a core of math that matters much more than other math in preparing you for the demands of career and college. And that column is a heavy depth of arithmetic understanding to fractions to the demand of linear equations by eighth grade. Going that column of readiness for algebra by eighth grade prepares kids. Think of math like the trunk of a tree or the handle of a fork. If you have that column, you can then do cool statistical stuff, because you can handle large data sets. You can then go on to advanced algebra, trigonometry, and then calculus. You can go on to applications of math, but without that core trunk, you are helpless in mathematics. Without number, you are nowhere. These core standards for the first time allow teachers to dare to focus, but then let's ask ourselves if the assessments do not then dare to focus. We have betrayed teachers utterly. Our curricular textbooks in grades, just to give you a sense of the numbers here, I was with Bill the other day, a typical textbook in a high performing country in math may be 150 pages fourth grade. Can you imagine the elegance and feel? I mean, it's hard to even imagine in your hands? What are publishers doing in response to the common core standards? Along with their 700 page book, they're now adding a common core portion. No. No, no, no. More expensive. Stupid. It's time to focus. It's time to give teachers the respect to do fewer things well that really matter. That's the great gift of these standards. Focus then allows coherence. Once you're focusing on fewer things, you can make far clearer the progressions between them. Math's beautiful thing is it's a logical system. The ordering of these standards if you want to really know it, it's not really about grades.

I mean, yes, it's third grade, fourth grade, fifth grade, but if you understand these standards deeper, they're building a logical structure, which is the logical structure of mathematics.

And finally, rigor. There are two types of people in math in my judgment. There are the kind of groovy, understanding people, and then there are the mean, rote people. And so, you know, which math is better when you understand or apply it or whether you can do it fluently. It turns out they're both wrong in the sense that around this smaller area of mathematics, it's equally crucial that you are fluent and fast about a core set of math facts, and I'll say facts, I'll say memory, I'll say practice. Yeah. At the same time, you understand what you're doing and can apply it quite broadly. That is only a realistic demand to place on teachers if we have the courage to do fewer things. Am I making sense? Because otherwise it's just too darn hard.

So that's the kind of promise in these standards. If we deliver that promise to teachers in assessment and if we deliver that promise in curriculum, then perhaps we can both have taken the skeptics seriously and made real movement in this country. Thank you so much for your attention.

So, Russ, the headline is you're a researcher I don't like, right? That's what the media coverage will be.

MR. WHITEHURST: Well, that was very, very interesting and entertaining presentation. Don't you think? I think it was. And it almost makes me forget my reservations. I've mentioned the people on the panel. Let me just say a couple of words more and associate them with positions on the stage. You can surely determine which of us is Nancy Grasmick.

MS. GRASMICK: I hope so.

MR. WHITEHURST: As David alluded to in his comments, Maryland is one of three or four states in the nation that demonstrated dramatic gains in student achievement over a 20 year period, a period that just happened to coincide with Nancy's tenure as superintendent of education for the state. She and I go a ways back, and it's very good to see her here. Jeff Nellhaus is responsible for one of the two consortia that's developing the assessments, and so I really expect that he'll be able to contribute to our conversation today in a number of ways, but certainly to give us a sense of what it looks like on the ground as these new assessments are being developed in terms of interims of their costs. Gerard Robinson not only has experience with assessments, but a lot of experience with the politics of assessments in Virginia and Florida. And let's make no mistake about this; this is not just a policy game. It's a political game, as well.

So let me start with what David doesn't like me for. And that's some research we published in the Brown Center a couple of years ago that took a look at the correlation between the quality of state standards and state performance on the national assessment of educational progress. And to make a longer story short, we found basically no association either with current scores or with gains over time. We look at states like Massachusetts that has excellent standards as rated by everybody and is the highest performing state in the nation on the most recent NAEP, and then we find states like California that also have excellent standards and are fourth from the bottom on NAEP. So in every quadrant of, you know, good standards, good outcomes, good standards, bad outcomes, you have an equal distribution of states.

So the question I would have, I'll ask David to start but ask other panelists to chime in, is why will it be different this time? Why is the construction of a new set of standards going to produce any more change than the standards that states

constructed on their own which, again, seem not to have been associated with outcomes?

MR. COLEMAN: You've heard enough from me, so I'm going to be very brief even though I'd love to really dig into this. First sentence, Jeff Nellhaus was the head of assessment in Massachusetts, not in California. That's a big deal, buys. The MCAS is one of the most beautiful exams we've yet been able to produce. Beautiful. I asked someone to look at assessments, an intern, to read over those assessments. I gave him no criteria which ones were good and, you know, without knowing the sheen around the MCAS, she found that the MCAS was the most beautiful assessment. Why? Because the literature passages in it were simply beautiful. So when you think about what teachers are going to look for in an exam, we may not know ourselves, but the MCAS has things like, and Sandra Stotsky who I sometimes debate with, one buttonholed me and said do you know we use the Wilbur translation of Molier on the assessment. And she's exactly right on this. It matters enormously. It does help the reliability of assessment to put shitty passages on tests. The use of commissioned passages on tests is ridiculous. Why? Because commissioned passage will never have the care for logical relationships that an edited published piece will. It just can't.

So whether it's informational text or literary, it's always a paltry imitation. And again, it's a model for the classroom. So I'm just trying to say, Russ, I think the deep point is standards in this country have not been nearly as muscular as they've needed to be in producing productive shifts in both curriculum assessments. You know, in the same research you provided, you kind of say what does show some power is curriculum of a high quality, and I don't misquote you, because I really want to get this exactly right because I want to be fair. And so what my interest is in, I think you can tell that I think

certain aspects of these standards really are a bit of breakthrough in their ability to focus more greatly than ones before them, in their ability to cause basic shifts that I fear you underestimated in your analysis. But the deeper point is the agreement which is what we're hell bent on is not leaving these standards an isolated force but trying to provoke productive improvements in curriculum and the measurement of efficacy of curriculum, in the quality of assessment in demonstrable ways that begin to bring work like the great work of Massachusetts, the great work of Maryland to many more states. So when you look at the PARRC consortia, isn't it cool that the Massachusetts guy who built this beautiful thing is now an architect for a broader -- I mean, I think that may be more important than all the efficiencies put together, that that talent, I really do believe this, I say this with an open heart, that the idea that beautiful craftsmen can make for a broader audience is the most beautiful efficiency of all. That's my view.

MR. WHITEHURST: Thanks, David.

MR. COLEMAN: Sorry, man. That was a bit of a hospital pass.

MR. NELLHAUS: I don't know how I can live up that reputation, but I would like to say a few words about, you know, since we're on the topic of building an assessment and how it relates to the standards. Just told the story when the Massachusetts standards were first released, we sent them to the schools in shrink wrap, and for the most part those standards sat on shelves in the shrink wrap for many, many, a year or two, before the assessment system came into play. Okay, now we have the assessments, people took off the shrink wrap and said wow, I think it's time to align our curriculum with these standards because they are actually going to be the basis for the assessments. So teachers started to look at the standards and, quite frankly, I think as many teachers as there were in the state, they all interpreted them a little bit differently.

Because, David, you think contracts are hard to read, standards are very hard to read, as well. And I challenge anyone to read the math standards and really understand what they're saying. They're technical, they're difficult to understand, and our teachers are not going to --

MR. COLEMAN: ELA literacy, different story, don't believe it.

MR. NELLHAUS: Our teachers are not going to understand the standards until they see how they're manifest in tasks that students need to do, and that's the role the assessments will play. They'll translate those standards into real tasks that students need to do, then those students' responses will be evaluated. And all of the sudden, we're laying on some values about what quality work looks like. So now things that teachers can understand come to light. So I think the power of MCAS was that it was very transparent, okay. So you can have a test, you can have the standards which are hard to understand, then you can have a test which is sort of a black box and produces a number. Or you can have a test that's transparent where the questions are released every year. The student work is released every year. How individual students did on individual questions is released every year. And now teachers have something where they can calibrate their own teaching to. Now they understand how these standards get manifest in tasks and when a student writes an essay, what is real good essay look like and why. So this is a very, very important role that the assessment system can play and that is the standards themselves will not move very far unless there's a way to kind of illustrate what we really mean by those standards.

So to the extent possible, we'd like to incorporate that in the new testing program. That's a cost issue, because what all of that, to develop new questions every year, to share all of this with teachers costs money. In Massachusetts, the median cost I

think, Matt's report was very accurate, I think most states on average are spending about $15 a student a test or about $30 altogether. Matt, MCAS costs more like $40 or $50 a test a student, $25 for the English test and $25 for the math test, but it was well worth it. Before MCAS was implemented Massachusetts was in the middle of the country on the SAT. Ten years later, Massachusetts not only had nearly the highest proportion of students taking the SAT, was also the highest performing state on the SAT. So there's something to say about teaching, I think there are external measure beyond the state test itself that showed between the SAT, NAEP, and other measures, that this really had an impact on student achievement. So I'll leave it at that.

MS. GRASMICK: So Russ mentioned that Maryland has done very well in terms of its progress on NAEP. We had an advantage. We decided that we wanted to inform our curriculum and our work by way of developing an innovated assessment as far back as 1991. And that as the famous MSPAP test, which really required students to integrate what they learned in terms of science, writing, they had to write passages, they had to give evidence, they had to justify their premise in their writing, and they had to do experiments and discuss why something worked, why it didn't work, et cetera. And we used the data, not so much for the student accountability, but to really use it as microscope to look at our curriculum and how we were presenting that curriculum to the students in various areas. It really required a lot of critical thinking on the part of the students, and I think as a result we lost something when we went to a multiple choice test that did not require writing. It required filling in the right circle, and I think that we had a built a foundation with our students early on because the assessment, of course, began in second grade, actually. And so those students had the benefit of the adjustments we made to curriculum.

We weren't as focused on our assessment in terms of No Child Left Behind as we were in terms of the development of a strong curriculum based on what we had done with MSPAP. And I think that is the success point of Maryland. We used that information diagnostically for students. It was particularly powerful with many of our minority populations. It was particularly powerful with our students who came from circumstances of real economic challenge. And it changed the way teachers taught. And teachers really bought into it. Teachers actually were trained to score the assessments. Most teachers went back to their classrooms and said after that experience, I will never teach the same way. It was a very powerful experience. I would like to say that while we were doing that simultaneously I happened to be a huge believer in AP. I think AP is elegant, and I am proud to say that Maryland has been the number one state in terms of the percentage of students achieving a three, four, five on the national tests in AP. And I think that we did that because we, again, wanted to set this very high standard for students. We wanted teaching to be done with fidelity to the intent of the curriculum for AP, and I think that's really contributed to Maryland's success.

MR. ROBINSON: Thomas Jefferson was actually the first board member and superintendent of schools here in D.C. Interestingly enough, he used that experiment in a way to help shape the educational landscape in Virginia. Florida is celebrating 500 years, part of it under a different flag, and when it decided to come on under the U.S. flag one of the questions that we had to address is how are we going to use money to adopt a new state into a federal system to support education in a different way. But what about results? Now I mention Florida and Virginia, because I've had a chance to work in two states. Two great states with a lot of results with very different models. When I was secretary of education in Virginia, I recommended to Governor Bob

McDonald that we not join Common Core. We know there are 45 states and three territories that are members. Virginia is not. Three simple reasons. Part of No Child Left Behind, Virginia in the early 90s began a conversation about assessments. And how are we going to use our money to asses Virginia kids on Virginia standards. That morphed into what we now have today which is Standards of Learning. When we looked at Standards of Learning we have tweaked it over the years, made some changes, and have had some good results. When it was time for the governors across the nation to decide whether to join, we decided not to join because the assessment that we had in place was good for what we were looking for. That isn't an anti-Common Core standard. It's more of a pro-Virginia. And if you've followed our history for a long period of time, we oftentimes move at a different drum beat and have stayed the course.

I moved to Florida, and Florida decided to become one of the 45 states to join Common Core. And for Florida it was the right decision. If you look at Florida over the last 15 years, there's very few states in the nation that have shown closing of the achievement gap with a state where 51 percent of the students qualify for free, reduced-price lunch, where 53 percent of the students are black and Hispanic, where you close the achievement gap and when you look at NAEP scores finding some of our students scoring at or better than middle class white students, and in AP a place like Miami having some of the highest scores in the nation for Hispanic students taking AP and seventh in the nation for African-Americans. Common Core made sense, because we wanted to join a consortium, I've had a chance to work with a few people here, to say it makes sense to have a common approach to a common problem and a common assessment.

And this came with politics, because we decided that it was going to cost money. And let me just give you the numbers. In 2012, we spent $16.9 billion, actually

$16.9 million; well really it was billion, $16.9 billion. Of that we ended up having a $59 million contract that was used for Pearson. Pearson's in charge of our FCAT 2.0 assessment. We had 4 million tests, 4.2 million actual tests that were scored and assessed for 2 billion students that cost the state $30.59 per student and $13.26 per test. That was 2012. If you look at 2009 to 2012, we spent an additional $2.55 per test and about $3.29 per student. So the big picture, economies of scale have worked out and we've been able to slow pace to ratchet up the amount of money we've invested in assessment. Now if you take the fact that we have $59 million and you look at the fact we did have a $16.2 billion budget, that's less than 1 percent. In fact, it's .003 percent of our total budget is spent on assessment. And yet there's a lot of righteous indignation, understandably, for too many tests. We have 180 hours in Florida for instruction, 8 hours is spent for exams. We have grading from three to ten in reading and three to eight in mathematics, and we also have end-of-course exams.

About a third of our school districts signed a petition, and ultimately the Florida School Board Association signed an anti-high stakes resolution, and they say we test too much. And I said if we test too much and we give too many tests compared to what? How did we find ourselves moving from the 40s up to one point of being fifth in nation based upon education week standards without having assessments, without testing in math and reading and science and now having end-of-course? Now I won't pretend that tests don't matter and that there isn't anxiety, but I also tell people there is anxiety with sex. And there hasn't been a lot of conversation about getting rid of that. Well, I figure I'll follow David's speech now with same talk.

(Laughter)

But in all seriousness, there's anxiety but the anxiety that drives me is

the fact that our children aren't competing solely with children in Massachusetts or in Maryland or in California or Texas. We have students in India, China, Africa, Brazil, South America, and they're going to have to compete in a world very different than our own. And for that reason, I think assessments make sense.

MR. WHITEHURST: So my take on what the four panelists have said is that you need good standards, but you also need great tests.

MS. GRASMICK: Exactly.

MR. COLEMAN: And you use them to inform correctly.

MR. WHITEHURST: And if you also have great standards and great tests together then it will be different.

MR. COLEMAN: And I mean, I think, the really interesting, what Nancy said that if found intriguing was the deep connection as in AP between assessment and curriculum design. So I think great tests may, just because especially the way people hear it, the real notion is great learning in the classroom, improving teacher practice, but also curriculum. I think is your deeper point, so --

MS. GRASMICK: Absolutely.

MR. COLEMAN: And it's that penetration, Russ, that I think we're trying for. I think. Is that fair?

MR. WHITEHURST: I think that is fair. We've heard end-of-course exams mentioned a few times today, including David mentioning that teachers really like AP exams, and they hate the state assessments. I wonder, again this is a question for the panelists, I wonder if there's a message in there, that the end of the course exams are better able to be aligned what teachers are actually trying to do in the classroom including the curriculum. And so the student's score at the end of the AP test or the end

of science class in 11th grade that has state end-of-course exam is seen by the teacher as a legitimate measurement outcome, whereas our high stakes end of the year test are not so aligned to curriculum. It is a bit of a black box or the teacher teaches what she's going to teach and then she drills the kids for four days on this other material they're going to be tested on the end of the year. Is there anything to that? And how's the Common Core going to then bridge the gap between this distant test and the actual material delivered in the classroom?

MR. NELLHAUS: I would just say I hope there isn't this disconnect between the standards and the curriculum and the tests. They should all be aligned. And so teachers who are having all the sudden to change their instructional program to prepare students for the tests didn't have the right instructional program in place in the first place if in fact the standards were good standards. So I think we need to get things better aligned, and there needs to be a lot of support for local school systems to actually implement curriculum that can support the standards. And the tests can play a role with that. I just want to say in terms of teachers or educators kind of, you know, hatred of assessments, it's not the assessments, it's all the accountability that's attached to them. So if you talk about AP, AP isn't being used to evaluate schools and districts. I mean it may be to a certain degree evaluating the teacher because you want to see if the teachers are getting their students to get a three or four on the AP exam. But generally they're not accountability tests, so as soon as you attach a lot of accountability stakes, if you will, to an assessment, that's what raises people's anxieties and their concerns about the assessments. And over the years we've attached more and more and more accountability to the tests, and that which costs were even better tests. Because, you know, where people are going to teach to the test because the accountability

consequences attached to them, that's creates a real burden on all of us to create very, very good tests because you want them to be worth teaching to. People are going to teach to them if they're high stakes attached to them.

MR. COLEMAN: I'm going to slightly differ and then move with you. I actually at the start, I've been obsessed with this question of why do AP teachers own this exam. I think what's fascinating is I think AP in our systems, Jeff, have been for most adults and children, must more clearly high stakes. They actually get you college credit so there are higher stakes for the kids.

MR. NELLHAUS: Higher stakes for the students.

MR. COLEMAN: Yeah, but, but, but I think in an AP course unlike really the, like who knows, the seventh grade teacher and the seventh grade math results. Do many parents walk up to the seventh grade teacher and say 23 out of your 30 kids passed the test? I've never heard of such, really, my kid failed the math test; it's all your fault. Like barely, right? Because that's seventh grade itself, but if you have an AP classroom and kids don't make it there are really, like this is actually observed. And so I actually want to open up a different possibility which is truly thoughtful assessment which really is even as, I know this might be crazy saying this, but more stakes for the children. Sometimes, see, I think it's kind of crappy; I'm going to say a controversial weird thing here. I think we really have to think of the stakes for children in our educational system differently. I think part of the problem with the teacher effectiveness movement is we've not thought hard enough about how to get kids and families equally into the game of feeling that theses exams matter, because it's a weird thing to have the job where you're being judged on progress, but the kids are kind of like, yeah who cares about the seventh grade test anyway. Right? And there are occasional visionary teachers, who are great at

motivating kids despite this, and they're highly effective, and I love them. But it's kind of a mismatch.

MR. WHITEHURST: I agree.

MR. COLEMAN: And in AP I think part of the beauty of is kids and teachers are in this game together, where they're like, yeah let's rock this thing. It's separately graded. So I just think we have to be very careful in a discussion of stakes, because sometimes more stakes, it's a kind of contradiction --

MR. NELLHAUS: We have the stakes, though.

MR. COLEMAN: -- can be more humane than, the only other thing I'd say about your point about courses is this, I think the fewer and focused in the standards is what may allow for a saner, so I think what happened, Russ, is typical state standards in math, as good as they might have been, very rarely focused as radically as these math standards do. So these core math standards in elementary school make it inescapable that K2 is the addition and subtraction of whole numbers, the quantities they measure, and third and fifth grade multiplication, division, fractions. It used to be that most math tests were surveys of a very large quantity of math topics. What that means is any curriculum could claim to be a lot. You see, you have this funny situation where the test is a Russian roulette system, so any curriculum, it's not like a course. So I do think the focus on the standards may take us one step closer to your hope of -- however, as head of the College Board, I'm thinking a lot about course design in earlier grades.

I want to just admit to you all. I'm thinking a lot about course design in middle school. I think middle school is lost. I think that kids need to make a lot more progress in middle school than they do today, and there needs to be much more, if this makes sense to people in this room, shapeliness to it. Sixth, seventh, eighth grade

science can't just be like whatever it might be. It has to have shape, a progress. It has to have courses you can be proud of. so you might think of the College Board as looking into, can we do course design in earlier grades that are, of course, aligned with the standards, but deeper than that. Give a trajectory. It has to be more flexible and modular so different states with different emphases in history or science can adjust them. It's a thoughtful design that allows for differences. But can there be enough of a core and a sense of a course? Because I think teachers have a deeper sense of identity, I can't explain this, it's a little mushy, when they feel they're taking their kids on a course of instruction. What does it mean to say I'm a sixth grade history teacher? That's not a lot. But if they say I teach a course that has this shape and my kids learn this stuff, and I can work with other teachers on that course, you've got a different environment.

MR. WHITEHURST: I'm going to ask the panel. I'm sorry. Go ahead.

MS. GRASMICK: I just wanted to make a comment that I'm so excited, and this is to David and Jeff. I'm so excited about the embedded supports that you're thinking of, because often we forget. For instance, in Maryland that 14 percent of our students are students with disabilities, and yet we have expectations for those students. These embedded supports are going to be enormously helpful to that population.

But I also want to say something about AP. When I began 20 years, 25 years ago, we only saw AP as for our most elite students. That’s it. I want to say that, as an example, our National Teacher of the Year, Michelle Shearer, who teaches AP chemistry, half of her students were students with disabilities, and every student achieved a four or a five on the national test. Isn't that incredible?

MR. COLEMAN: Incredible.

MS. GRASMICK: We can do it. We can do it.

MR. COLEMAN: That's amazing. Beautiful.

MR. ROBINSON: In Virginia and Florida, the creation of standards is not done absent of teacher involvement. There are actually committees that exist where they actually design it. The rub comes in when it's time to link accountability to what is taught, because often those decisions are made in state graft. It takes place in the political legislature. It takes place with lobbyists. And that's where the some of the rub will come in. I'm of the belief that we need an accountability system, we invest billions of dollars into our education system, to have a better idea of are we getting the best bang for our buck and return. It will mean that some teachers will receive financial rewards, as we have now an upcoming merit pay program in Florida, a merit pay pilot in Virginia. But it will also mean that there are some teachers who will lose their jobs because the results aren't there. This isn't anti-teacher stance. It's just looking at accountability for students today for an economy really that expects a lot more accountability than we've been getting so far.

MR. WHITEHURST: I'm going to ask the panel on more questions. I'll ask them to respond briefly because I want to make sure to give you an opportunity to ask questions from the audience. The question is one we discussed in our green room before we came in here this morning, and it's about sustainability. one of the fact, I think, that came out today, both in Matt's report and then the discussion of the panel is that the Common Core is going to cost more than what we're spending now, still a small amount with respect to overall educational expenditure but more. It's also technically quite challenging to launch, and it's going to need technical oversight going forward. So the issue is one of sustainability. We've had a big influence of foundation and federal money to support the effort. What's going to keep it going five years from now or three years

now when states are entirely on their own resources?

MR. NELLHAUS: Well, sustainability is a huge issue for PARCC right now, and we have the support of the Council of Chief State School Officers has just contracted with McKenzie and Company to actually work with both consortia to think about sustainability. But I think we're laying the groundwork for sustainability now in that we're working very closely with the leadership in each state, each of the states to design the program. This isn't about Achieve playing a project management role and putting the assessments together, but we're meeting regularly with leadership in all 23 states to make all sorts of decisions from high level policy decisions to very operational kinds of decisions with the assessment directors in the various states. Without doing that, without getting that buy-in, I don't think the program will sustainable. Some of the issues around sustainability also have to do with procurements in the future, and this is a big issue right now for PARCC and I think for the two consortia, and that is how are we going to purchase the services we need. And you need to identify a state that will be a fiscal agent for that. Leadership in those states can change and that can affect the whole consortia. So there's a lot of issues around sustainability, but it's not an issue that's asleep right now. We're trying to tackle it head on.

MS. GRASMICK: I would just applaud the efforts that are being made, but in my opinion, we have a changing cast of leadership in the states, and people who are in Chief State School Officers can buy into it, but then they change, and we've already seen some of that happen. And so I think there needs to be an identified sort of national entity but not the federal government entity, somehow that’s put together by the states that can ensure a level of consistency and sustainability. And so I don’t think it's going to work otherwise.

MR. ROBINSON: It's marketing. We've got to do a great job of letting people, for example in Florida, know what Common Core is, what the objectives are, and what role they can play. And so over the summer we had a four day workshop where we had 8,000 educators participate to go through always from A's to Z of Common Core and what it will mean. But more importantly we, with some support money from the Gates Foundation, had a campaign where we had focus groups with parents, with teachers, business leaders, real estate agents, very important in Florida, to have a conversation not only about Common Core but also the change in assessments we have for our A to F grading system and other things. Interesting that in the state where you've had so much accountability, so many standards, there were people who were asking kind of like one-on-one questions. Well, what does this really mean for my child? And if the school moves from an A to a B, what will that mean? And are we getting rid of FCAT if we're adopting Common Core? What about science? You know, it doesn't cover science. But it gave us a really good chance to talk about something very simple, education and what it means. And so marketing is going to be important because the political branch will change with time, governors will come and go depending upon leadership, same thing with state board members, but the citizens where you have children in this classroom or not, you're the ones who are investing your money in it. We have to do a better job of marketing, and we've at least started that process before.

MR. COLEMAN: To be brief, excellence in this world is rather rarely achieved. It's a fragile and beautiful thing when something is a good thing. And so to steal an idea from Russ, I wonder where there’s a version of NAGB that has a state-led component. The NAEP is, in my mind, a remarkable achievement of thoughtful research, careful design, principled execution that is within our midst. And so if there's any way to

take the federal taint away, to see this as truly -- we must sustain the state-led beauty of these Common Core standards. As we sit in Washington, let us not underestimate the dysfunction of this town. And the beautiful thin in the Common Core standards is that at the most partisan time in this country, 46 governors got together and did the right thing with their Chief States. It's like a fairy tale, right? An imaginary country in which Democrats and Republicans work together to solve problems. So please do not underestimate the principal and seriousness of state-led that I can sit next to the foolish leader of Virginia who kept with the Virginia standards, but I utterly respect his right to make that terrible decision for his state.

(Laughter)

And that's a deep thing, guys. So in whatever sustainability model we do, we must celebrate and strengthen state leadership as a principal even if the leaders of it shift. Because governors do, too.

MS. GRASMICK: Yes.

MR. COLEMAN: But it's okay, because the notion of state leadership is a deeper one than the individual's leading it. The second thing I'll say about this is the College Board is a long-standing institution, and we will do anything in our power to cultivate sustainable models of state-led leadership so I want to be rather clear. We have no ambition to control or run these things, but if we can help states build kinds of institutions you're describing, if we are in any position to cultivate or strengthen such institutions, we are utterly committed to it and we're not going anywhere.

MR. WHITEHURST: It's time for questions from the audience. I ask you to raise your hand. I will call on you. Someone will bring you a microphone. Tell us who you are. Make the question brief, and I'll try to get the panelists to be brief in their

response, as well. Questions? Right next to you there. Thank you.

MS. HANSEL: Hi. Thank you. This has been a great panel. I'm Lisa Hansel. I'm the editor of AFT's magazine, American Educator. And I'm just wondering -- this is for all the panelists -- to what extent are you working with traditional or alternative teacher preparation programs? To me, one of the great things about the Common Core is that we finally have common content that could be the basis of teacher preparation. And, you know, David Cohen has talked about this, that one of the big problems with teacher prep is this sort of lack of content. You're being generically prepared to teach anything. The job's harder than that. You need something specific to bite into. And also bringing this to the testing side, helping our teachers know a lot more on day one about how to use the information from tests in that diagnostic way, not having to let years go by before they start to grasp that, which is also a very, very complicated thing. So thanks. Any information should be great.

MR. COLEMAN: Can I? You know, Russ, there's another illusion I made need you to break to further discourage me in my professional life, which is there's in AP it seems that professional development may be effective, but I'm so skeptical of this, because the research on the efficacy of professional development is endless tragedy, right, so a massive investment with zero result. Terrifying. So I really, I need help here. Is it true? Because we have a perfect example of it in AP, that training teachers not to like to teach, but to teach a well-designed course helps them improve their practice. This is a studyable question, and I really wish, this is rumored to be true, but I will admit, I'm a skeptic. It's my own institution. I believe it from a lot I hear. I do, too. But I want to know it. So I think that's one question, is can we demonstrate that we can invest with confidence, that giving shape to teacher preparation around specificity

course knowledge produces results. And at the same time, can we use assessment artfully? So that people who want to be teachers or are beginning as teachers demonstrate a command of the knowledge, rather than these vague assessments we have. We at the College Board are thinking a lot about whether we have a productive role to play in this discussion, but we're very uncertain today. We don't play much of a role in it, but we're wondering about what AP teaches us about professional development that's embedded more in a notion of a course of preparation of teachers.

MS. GRASMICK: My comments are aspirational. I would hope, I would hope that we can really confirm our belief, as we see with AP, that in really having a solid curriculum and having people who have subject matter expertise that we can change the way we prepare our teachers. I really think it's the hope and, like David, I don't have the evidence, but certainly from everything I've observed that is true with AP. And the sense of pride in the delivery of that curriculum.

MR. NELLHAUS: It's real stuff.

MS. GRASMICK: It's real. The second thing I would say is that alternative teacher preparation, obviously, in most states we have some cohorts of alternative preparation, and so I don't think it's going to end, but I don't think it's going to overtake more traditional preparation. I just hope that preparation changes. But we have not achieved anything in my opinion with assessment if we cannot use that assessment diagnostically in terms of intervention for that individual student.

MR. ROBINSON: So we have traditional colleges of education for certification in Florida. We have partnerships with these local school districts and other organizations certified at the local level. We also have, I think, it's the American Board for Teacher Certification Education provided. They're in Florida. In fact, Tony Colon is one

of their board members. He spent a number of years here in Washington, D.C. One of the projects that was put in place before I left was to really identify where the achievement is, where the challenges are at school, and to identify the teachers and what school he or she attended or where they received certification. It's not secret. We kind of have an idea already who's doing it, but being able to identify quantitatively is going to be important to identify what schools are doing right and what schools need to change.

MR. WHITEHURST: Next question. Here.

MS. HUTA: Thank you. Good morning. My Joanna Huta and I work in the Department of Education.

MR. COLEMAN: I'm sorry. Where do you work? I didn't hear you.

MS. HUTA: Department of Education.

MR. COLEMAN: Okay.

MS. HUTA: There was a comment earlier about teachers saying that they were doing too much testing and so on. And I think the major complaint there is interrupted instructional time. It's not too much testing, but the time that it takes away from instruction. So I would like to hear any comments about embedded formulative assessments that would not take away time from instruction and whether anybody sees future in that.

MR. COLEMAN: Can I just say one thing about that? I am convinced and again I'm really turning to the two of you, but there's very interesting research that assessments themselves are instructive. I think this is one of the stupidest bullshit distinctions we've got, that I'm taking a test, but I'm not learning. I'm sorry, even formative summative, like I'm daring to like, that is, high quality work in an assessment environment I think has been shown, but again it should held very, I think there's some

interesting work on this. So I think that this dichotomy is killing us. So what I mean by that is I'm kind of trying to agree with you, that whether it's a formative assessment or a high quality summative assessment, the practice of doing something demanding is itself an act of instruction. It does not subtract from it. This is madness. And we've got to stop, like, giving into this talk. We've got to stop flattering this talk. Why? Because, guys, this is so stupid. You get this rage up that we're wasting time testing and make tests shorter and shittier. And then people say the test is all multiple choice and bad. And we've worked ourselves into a set of stupidities when really we, I'll say it: we might want longer, more thoughtful exams. AP is not a short, shitty test.

MS. GRASMICK: No, it isn't.

MR. COLEMAN: It has more regard from teachers, because it's work worth doing, guys. That's the question and yes, but I'm saying and that's high stakes. And so we've got to even own, because you know, it's like if it's formative and wastes time it's okay, but if it's not -- but I can I tell you? There's a lot of shitty formative assessment stuff out there. Those interim assessments that many states and districts are using right now that are low stakes. Oh, my god.

MS. GRASMICK: They're horrible. Horrible.

MR. COLEMAN: In terms of quality. So I think it's like the real game here is quality, but as a political and social group, we've got to stop indulging in these statements. They're very costly, takes away from instructional time. Jeff is in a very dangerous situation here where if he makes the test long and beautiful, he gets this whipshaw attack, right. And then you're stuck. You see what I mean? We've got to change the game, guys --

MR. NELLHAUS: Already.

MR. COLEMAN: -- if we're going to shape this country.

MR. NELLHAUS: I just want to say that technology's going to play a bit role in this in the future. Right now a lot of the testing kind of disrupts that curriculum now because of the logistics of administrating these tests.

MR. COLEMAN: That's exactly right. That's a very important thing.

MR. NELLHAUS: The boxes come in. They've got to be unpacked. People go to get there. The students with disabilities rightfully need accommodations and they need adults to help them with those accommodations, so Nancy talked about embedded supports before. What she was talking about is that on an online assessment, a lot of these accommodations will be built in online. If you need the test read aloud, you'll have a pair of earphones. There will be a text-to-speech accommodation built right into the assessment. So we're looking at ways to make this much less disruptive, and I mean, I think the vision is that in the future student will come to school just like you and I come to work, and you'll have a computer on your desk. And once that happens, we'll be able to do this much more efficiently. It can become more part of the instructional program than it currently is. I agree with David. A good test will be instructional but there are little --

MR. COLEMAN: Yes, I totally agree with you. Getting rid of all this crap around it. Brilliant, brilliant. Exactly right.

MR. NELLHAUS: Right, so I think the technology as we move forward, the vision here is going to look a lot differently --

MR. COLEMAN: Couldn't agree more.

MR. NELLHAUS: -- because the technology's going to be in place to

help this.

MR. COLEMAN: Totally agree, Jeff.

MR. WHITEHURST: Here. I'm calling on people who had the foresight to sit on the aisle just because you're easy to get to. Let that be a lesson to you when you come from the next one.

MS. CLINE: I'm Indor Cline. I'm a development consultant working with non-profits that work within the school districts. My question is two-fold, one with regard to return of investment. I'm particularly interested with regard to Florida. I know that the onset, it's a big cost up front and then you see the returns with lessened costs in the budget. What are you doing with regard to creating buy-ins with partnership, particularly with the business community and the non-profit community with regard to sustainability, number one? And number two with regard to marketing, how are you using social media, and I don't mean just Twitter, a comprehensive social media program that includes PSAs and the use of texting?

MR. ROBINSON: Someone here's got to leave in about two minutes, so I'm going to give you a 30 second version and can stay here afterwards to go into detail. I'll work backward. When we decided to release our new scores and we knew a number of schools have A's and B's, I partnered with four or five superintendents, did a PSA. They used their technology, and they went to over 100,000 homes individual districts, first time I had been done. I would also make good use, I mean there's always YouTube and Twitter, but to me the concerted effort to make sure that happened. I also was involved in some Tweeting, but that's another part. In terms of non-profit partnerships, we actually have a partnership with College Board so that all of our students in Florida have access to AP tests free of charge, something that we do and we have in place. We partners with

the Council of 100, and we actually partnered with them to make sure that our codes in business preparation courses were in line with industry numbers, because their numbers are changing real time, some of ours were not. So a student could have been taking courses to prepare him or her for an IT job, the identification number changed in the business sector, six, seven months ago we were teaching on A, they were doing B, so that was a collaboration.

With the Chamber of Commerce, they're actively involved in the work that we're doing. They also play a role in saying why Common Core matters. In fact, we were here in Washington, D.C. A grant from the GE Foundation gave an opportunity for us to have a conversation about Common Core and what it means for Florida and business. And more importantly, Common Core allows us to work with higher education partners in ways we haven't before, because remember colleges receive our students. They're some of the ones who are saying they're not prepared. Get away from the tests. They're just saying we're seeing everything here. So there are some examples.

MR. WHITEHURST: Well, thank you very much for attending the event today. And I hope you will join me in thanking the panelists for their presentation.

(Applause)

MR. COLEMAN: Thanks, guys. I have to run to the train.

* * * * *

CERTIFICATE OF NOTARY PUBLIC

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

Carleton J. Anderson, III

(Signature and Seal on File)

Notary Public in and for the Commonwealth of Virginia

Commission No. 351998

Expires: November 30, 2012