THE BROOKINGS INSTITUTION

PROMOTING K-12 EDUCATION TO ADVANCE STUDENT ACHIEVEMENT

Washington, D.C.
Tuesday, September 27, 2011

PARTICIPANTS:

PANEL III: NEW ASSESSMENTS FOR IMPROVED ACCOUNTABILITY

Author:

DEREK NEAL
Professor of Economics
University of Chicago

Moderator:

MICHAEL GREENSTONE
Director, The Hamilton Project
Senior Fellow, The Brookings Institution

Discussants:

PETER GORMAN
Senior Vice President, Education Services, News Corporation
Former Superintendent of Charlotte Mecklenburg Schools

MICHAEL MULGREW
President, United Federation of Teachers

ROBERT L. HUGHES
President, New Visions for Public Schools

* * * * *

P R O C E E D I N G S

MR. GREENSTONE: Okay, we have our third panel. Derek -- we are very fortunate to have Derek Neal, who is a professor at the University of Chicago, a former colleague of mine, current friend. And he's going to talk about new assessments for improved accountability.

MR. NEAL: Okay. So I think one way to understand what I have to say today is that Roland was correct when he said that a lot of the incentive systems that have been used for educators haven't worked. And what I want to talk about in broad terms is the idea that the vast majority of incentive systems that have been used in education, if you took them to a professor at the Sloan School or the Harvard Business School or the Booth School at the University of Chicago that teaches personnel, economics, and incentive design and contract design, and you asked them would this work? They would have told you no before you started.

And so one of the things that you want to think about when I go through what I have to say today is, I'm talking about taking seriously the task of designing incentive systems for teachers that have a chance to work. I'm not confident that if you did things the way that I described you would necessarily get wonderful results. But I think it's time for policymakers to either take seriously the issue of designing incentive systems in ways that make sense from what economists know about how incentive systems work, or abandon the notion of assessment-based accountability altogether. That what we've done so far is really try to do things on the cheap in a very haphazard way.

And the first aspect of doing things on the cheap is that we use two tests -- we use one test for two different purposes. So in most accountability systems or performance-pay systems, you have a test that's being used to measure how the students are performing in a system over time. So you want to compare the 5th grade scores in 2005 to the 5th grade scores in 2010. And at the same time, you're taking those test scores and creating performance metrics for the educators.

However, the properties that you want in a system that's going to be creating consistent information about student achievement over time are not the properties you want in a performance metric system for incentives. The thing you want for consistent measures of student achievement is, you want enough predictability and overlap and common formats that you can rest assured that you can score the

2005 test the same way you scored the 2010 test.

But if there's that much repetition and that much predictability, the best response of the educators is to actually have test prep sessions and memorize answers, rather than teach. If you have, on the other hand, I want to start with a system where I'm going to design it to create performance metrics for educators, what you would want is to have a new format every time. You would want it always to cover the curriculum, but with different types of questions, different formats, completely unpredictable in any kind of coaching way. But then if you did that, there'd be no way to put the scales together, okay?

So you can't come up with one type of assessment that's going to serve both functions. And when you try to do that, what happens is that the predictability that makes the scaling possible in theory leads to the coaching and practice that corrupts the scale. Okay?

And the most important point about this is that there's actually huge losses that aren't measured in the data. If you have a lot of creative activities that are being crowded out that build true subject mastery, that's a big loss for the kids that's not captured in the data. Okay?

Now, there's a paper in empirical sociology, it's 30 years old now. Where a guy named Campbell who was at Northwestern reviewed a bunch of case studies of different government agencies that took some statistic that they had been keeping just to monitor how they were doing. It was a performance thing, a consistent scale thing.

They took it and said, okay. I know what we'll do. We'll turn this into an incentive device. And what happened in every situation that Campbell reviewed -- and then since then in Heckland's work on job training and all of my reviews on education and in a paper by Rothstein on many different government agencies -- the performance measure gets corrupted by the gaming behaviors, and the gaming behaviors hurt the citizens that are supposed to be served. And ex post, the performance statistic doesn't measure what it's supposed to measure anymore, okay?

And so this is not just about education. This happens in every realm of government when you take a statistic that wasn't designed for use in an incentive system and say, look. We can have accountability and incentive on the cheap. We'll ex post, just attach money to the statistic. Okay?

And Race to the Top is not going to get you around this, okay? It might get you better coaching because the exams will have other than multiple choice questions. But so does the CPA exam.

The CPA exam has demonstration of its et cetera, but it's on one of these IRT systems that are supposed to keep a consistent scale, so they repeat a lot of questions and keep the same format for the computer simulation.

And when they went to this IRT-type format in 2005, in just 5 years the pass rate went from 39 percent to over 50 percent. That was not the improvement in accounting education in America. If you look at the exam -- if you look at the websites that offer these courses, you know what it is. Over time, these places where you pay to learn how to take the exam developed simulations that looked exactly like the exam. And you practice those test taking skills. In 2011, they changed the computer format, okay? And it dropped dramatically.

So -- I have to go much faster. (Laughter) You have to design the first step is to have two tests, and you have to design them to avoid coaching. This is a new project we researched.

The last thing is, you don't have to have scales to have incentives. You can put the kids in groups with their peers, rank them at the end of the year according to how they're doing relative to their peers, and that is -- that percentile score is a fraction of the kids in California like me that I beat. You take that number and average it over all the kids in a grade in a school that are taking, say, fifth grade math. That's a winning percentage for that fifth grade math team. And I have a paper that shows how you can use that winning percentage to -- and contest and tournament theory and economics to build sensible accountability practices and incentive systems around that winning percentage. (Laughter)

I'm not hooked up. Oh, there. All right.

For me, that was pretty good, Michael. (Applause) I was close.

MR. GREENSTONE: I was impressed. (Laughter) We had put in a little extra time in the session, but now we're going to take that back.

Okay. So one thing that we really pride ourselves at the Hamilton Project on is taking just super-fabulous academic ideas and trying to make them -- and I can say this is not about Derek's -- this is about my own academic ideas. They're misshapen, they're ill-formed, there's an arm sticking out here, there's an arm sticking out there. In a way that I have sadly found out, people in Washington and in government in general find disconcerting and not very useful.

And so I think the question I want to ask our panelists is, I think Derek really has a fantastic idea.

What can we do about it to turn it into something that could be implemented in a way that would be used to a new method to reward -- assess teachers and reward achievement?

Bob, do you want to try to take a stab at that?

MR. HUGHES: Derek's presentation --

MR. GREENSTONE: Remember, you got off -- you didn't have to talk about the Secretary of Education --

MR. HUGHES: Yeah, I didn't get the last one. I understand that, and now I'm being punished.

MR. GREENSTONE: Yeah. (Laughter)

MR. HUGHES: I'm still struggling with the concept, to be honest with you. I made the right decision becoming a lawyer and not an economist.

But as I understand it, in the system you're doing a tournament style where you're basically ranking kids based on how other kids do, their peers. And you're dropping the scale score and using that ranking to make judgments about teachers. Is that right?

MR. NEAL: The big features is -- and this goes to the New York City bonus program that Roland talked about.

I wrote this survey paper surveying all the studies on teacher incentive systems. And there are only three that produce decent results. And one of them was one in India where they brought in a new test that nobody had seen before. The other two were in Israel and they were real tournaments where there was a fixed number of winners and you weren't competing against a standard.

And one of the things, if you look into the details of this New York City bonus program, 89 percent of the people won. And they won in a year where it appears that there was some funny business going on with the scaling of the state assessment.

And so if you wrote down a standard contest model, you would predict that that wouldn't have big effects in terms of output because this is a contest that everybody knows they can win without putting forth a lot of extra effort. And so what I wanted to do was to have a system that uses only the ordinal information and it's a real contest. For every implicit student comparison in PS 105 to a kid in PS 107, there's -- one is going to win, one is going to lose. And so every dollar of bonus money that the

teachers in this school win is a dollar of bonus money that somebody somewhere else lost.

MR. HUGHES: Got it.

MR. NEAL: And that's the other main feature of incentive design that you see in business schools that you never see in education, which is you use relative performance measures in environments where you're worried that if you tried to set targets, those targets could be manipulated. And because psychometric scales are so fuzzy, targets can always be manipulated. And so having true contests where there's always a winner and a loser is a discipline device that is almost never brought into accountability and incentive --

MR. HUGHES: Got it.

MR. NEAL: -- pay in education.

MR. HUGHES: So, you know, consistent with the last panel I think it would be interesting to do models and small tests of change to see how it actually plays out and correlates with what we know about good teaching.

But even at the description that you made, I think there are a set of questions that are problematic. And we kind of know this based on the New York experience, which has been mixed.

One is, the statement comparing peer-to-peer is much more complicated in a system as diverse as New York City. We have a peer horizon in our current assessment system. What we found is those peer horizons don't actually capture a set of intangibles, like admissions policies in schools or the mobility of students in a particular school. Or the absence, frankly, of entering test scores. So that getting that peer-to-peer is very complicated to do.

Secondly, depending on how sophisticated you get in your ordinal ranking system, it seems to me you're never going to be sure if you have enough of any particular type of student to make it meaningful.

MR. NEAL: So, 10 states are already doing this implicitly. Because every place that uses -- like Colorado that uses a student growth percentile software to come up with a student -- every time you see the word "student growth percentile", that's one of the scores that I'm talking about.

Now I've been working -- I've got a beta version of it. I've been working on my own version of that software to try to refine it and get better estimates. But there are 10 states out there that

are already trying to estimate if we take a rich set of the student's prior achievement and use that as conditioning information, what's the distribution of scores that students with those characteristics are drawing from? And then you take an actual score and put it in that distribution and you get a percentile.

So, the measurement aspect is not anything new. Everybody that uses that Bet and Beaner software in those 10 states is already doing this. I'm trying to look at ways to actually improve that software and do it better.

But in terms of the practicalities of estimating these percentiles, it's going on already.

MR. HUGHES: Okay.

MR. GREENSTONE: Michael, I think one thing that there's been a big concern about with accountability is that teachers feel pressured to teach the test. And that might not be the best --

MR. MULGREW: His research clearly shows that, yes.

MR. GREENSTONE: So what would be a more holistic way that teachers would recommend to judge teachers, I guess?

MR. MULGREW: Well, even on the New York City whole school bonus, when they researched their findings -- and yes, one year 89 percent received a bonus -- the teachers at that time did not know that the scales on the test were being -- let's say this correctly. Were being moved in a direction that would show greater achievement. (Laughter) The teachers had no idea.

This is not the teachers making these decisions. But the findings, when they studied it, they clearly said the teachers were doing the work. They were not incentivized to do more work because they were in an incentive program.

So we're getting down to -- what you're proposing is that we start everything as a straight-up competition between peer groups. Well, if the teachers who have been involved with incentive programs -- and we all -- the research is clear that these have not worked. But the one finding we keep coming up with is that teachers are saying this is not what incentivizes me. More money. That is not the reason, that is not something that is moving me in the profession which I have entered into.

So when you look at accountability you have to look at a multiple measure system. When I -- we look at the countries who we always speak of -- the Singapore, the Finland -- they have -- their evaluation processes are professional growth systems. They're not evaluations. It's framed upon

being a professional growth system from the first day you enter the profession to the very last. You should always be growing.

That is not the approach here in most places in the United States.

MR. GREENSTONE: Okay, so can I push back on that for one second? So the thing that I am hearing when you say that is, but on the other hand we know that getting a masters of education doesn't lead to more effective teaching.

MR. MULGREW: Correct. So that is why when I said before your first day in the class is usually not your best day by any means. And we've spoken about already about the preparing someone. If you have a teacher accountability system that is -- the framework is first and foremost, how are we moving, helping teachers move constantly in the profession of pedagogy?

But at the same time, take -- if we have that type of system, I know there are places. We have New Haven, who has very interesting findings after a year or two. You have Montgomery County, Maryland who has used a professional growth system. There are very interesting things going on and we're getting positive results. So, I think that's where it has to move in terms of accountability.

You also have to have a fast, fair process for removing teachers who do not do their job well. I was a mentor. I was a union leader. And I clearly, at times, I would have to tell people, listen. I know you want to help kids, it's great that you want to do that. But this profession is not for you. And just make it fast and fair.

MR. NEAL: So, I want to be clear here. I didn't have time in seven minutes, but I've written in many places that these percentile performance indices based on test scores can only be one small component of a larger system. Because if you only pay attention to the achievement, you won't get attention to the arts, to music, to social development, to the other things that we all want to go on. And I've also written that you want to somehow bring things like school inspections or the Cincinnati Teacher Evaluation System into it --

MR. MULGREW: Very interesting stuff, right.

MR. NEAL: Okay? But the point is -- where we probably part ways is that I want all of those to be on this same relative performance criterion. And so that you end up with a performance -- these set of performance indices where everybody is ranked relative to each other and there's always

going to be some level of relative performance where if a team of teachers or school falls below that, we come in and we, you know, deny public funding, reorganize -- that there's a distance between the median performing school and how you're doing. That if that gets big enough, you lose your right to receive public support as a school or as a team of teachers.

And I think that's a very important discipline because then the standards move endogenously. If there are new developments in pedagogy, if people are figuring out how to do things better, then that median point in the distribution moves with it. And you don't have to come up with some new projected standard, and you don't have to worry about somebody in the psychometrics division at the State Department of Ed monkeying with the scales.

And so, teacher incentive systems have not worked well, but they haven't really been seriously designed to work well. Because they haven't been based on assessments designed for use in incentives, and they haven't been --

MR. GREENSTONE: Well, hold on here.

MR. NEAL: -- true contexts.

MR. GREENSTONE: Now we are very fortunate to have someone here who has, I believe, set up a system. At least partially -- part of the compensation was based on assessment. Would you want to talk a little bit about that, Pete?

MR. GORMAN: Well, first of all Charlotte is nowhere near there. There is so much more work to be done, and I'm a heck of a lot smarter now that I'm not there then I was when I was there. It's remarkable how that happens.

I like the fact that this proposal takes out the within-school non-random assignment of students to teachers. I like that piece. That it goes and looks at things in a larger aggregate. And I think that piece is incredibly powerful.

But the nuance is, what is the right amount? What is the piece that it plays? Because you've clearly said, it's not the whole thing, it's one component. What are all the other components?

For example, we're trying to figure out in Charlotte of what is -- how do we measure in an evaluation the impact of teachers on other teachers? What do they contribute to a professional learning community that makes other people better teachers in how they share? And then how do you measure

that? Because that's a huge piece we're finding out that our kids who grow the most in Algebra 1 have teachers who work as a team, share, trade kids back and forth through the course of a day, and do those things. So how do you nuance and measure that?

And I really made a huge mistake when I was in Charlotte. I came out too quick and too early on value added. And then once you put that out you can't go back. So that's why I think the system is going to get better now that I'm gone. Because that can be tied to me as a person, and then they can go now and fix those things. And I think that's powerful.

But we use student learning objectives. We added a component that was student growth. And I like this example, though, at looking at a student growth model that is more towards a format such as this. The question is, though, should there be a piece that's individual, that's team of teachers, that's school? Should you get that nuanced? Do you get so far along the line that you have so many measures that no one measure is impactful or meaningful any longer?

And what I'm learning is, I don't like the current system. We're trying to do something that is better and we're experimenting with different options. And the biggest mistake we'll make is if we tell folks we've now got it, here's what it is.

MR. HUGHES: I think -- yeah, just to build on that. I think we've all pushed hard kind of for some sort of empirical basis for making judgments about teachers. The problem is as I understand it, we end up with curves that don't tell us how to move the middle. And we've got a few saints out there in the education reform movement and we've got a lot of sinners. And the challenge is to move the middle and make them better.

So in some ways I'm saying what Michael said. We've got to figure out evaluation systems that give teachers actionable information on how they improve their performance as much as we need to figure out how to make decisions on who stays and who goes. And we've spent an enormous amount of energy in this country proving that we can demonstrate the tale statistically.
But we also have correlational evidence that says principles know who those teachers are. And so the question I think is, what are the systems that are going to move that middle to create vastly larger numbers of good teacher?

And you're part of that, and I think the multiple measure question is the crucial version.

MR. GORMAN: So it's got to be a teacher development and evaluation system? Not an evaluation system standing on its own.

MR. HUGHES: Exactly, right.

MR. MULGREW: When the -- where you were speaking before about looking at whole schools, which is why we tried in New York City the whole school approach, I think in your system -- because as you pointed out in your presentation and everybody here is agreeing, the current systems that we have on the table for doing this are just -- they're not working. We can go into the scandals, everything else. Be teaching to the tests are just bad enough on its own. Looking at whole schools it's a little bit more -- I think there's a lot of value there to look at to see how whole schools are performing. I don't know if this would ever work on an individual level, but I think the approach and coming at it from your point of view is interesting. And at least -- it's changing what we're talking about.

And it's -- like right now, what concerns me is there needs to be a sharing. If someone is doing it right, we have to make sure we set up a system where that information will be shared. Not, oh, I'm doing it right, I want to keep my bonuses. You're not going to learn what I'm doing. We don't want to go to that.

MR. NEAL: I'm very opposed to ever using this to give scores for individual teachers. Because I want the situation where if Ms. Smith is a master teacher across the hall and knows Ms. Jones is screwing up, that she walks across and helps her.

MR. MULGREW: That's --

MR. NEAL: And it is insane -- and I get very annoyed with some of my colleagues in the economics of ed literature that pretend that if we put really high stakes on individual measures that we wouldn't weaken that cooperation. I mean, I think everything we know about how people respond to true zero sum games like I've proposed is that you don't want to use those among people who are supposed to be helping each other.

But I don’t think there's anything wrong with PS 105 competing against PS 106.

MR. MULGREW: I don't -- I agree with you there. Right -- I was in Williamsburg two weeks ago in a school in New York City because it's a school I know who does very well with a tough population. And I've seen their enrollment trends over the last two years increase significantly in the two hardest populations to educate, which are newcomer ELL students and children with special needs.

And I get concerned quickly when I see that happening. Is it an enrollment issue actually that the school should be dealing with? Is someone from outside the system? New York City, we think of a lot of conspiracies a lot of times. (Laughter)

So, I went --

MR. NEAL: Everybody is so honest in Chicago, we don't worry about it.

MR. MULGREW: Yeah, I know. (Laughter) So, I went to the school and it's a school that any teacher would want to be in. And it's a school that any parent would want their child to go to. But there was this significant increase. And I asked them -- and it's a great staff, because it's always the principles and the teachers talking to me together and pushing at me.

And they said, no, no, no, no. We're recruiting these students. I've never heard that answer before. I said, why are you recruiting them? They said, all of the other principles are closing programs around us so they don't have to take these kids because they know they're the hardest kids to educate. And they're afraid that their grade, which is a competition piece, will drop. So they said, that's not why we got involved with education.

So, I've never been in a small elementary school where they had four self-contained ELL special education classes. Because they went out and recruited those kids to go to the school from the neighborhood. And that's the work -- see, that's what I'm always afraid of. If -- when I have concrete evidence of that. But it's anecdotal only. But when you see that that's how people react to this, that's scary stuff. That is very scary.

And how we get accountability right and also deal with that type of issue, that's what I hear you're really pushing at trying to get to that. And that's why I thought your paper was so interesting to me. You're pushing at it, but that's the realities we're dealing with in the schools and in the communities. That there are schools who are now saying, no. Those kids are very difficult. And I would love to do it, except my school's very existence is based upon my grade and I know if I don't have those kids I have a better chance at a better grade. Because the weightings are still nowhere near nuanced enough to deal with it.

MR. NEAL: But in mine, you only compete against the other schools that are serving those kids. So that's not a disincentive.

MR. MULGREW: But on the special education side, we've been dealing with this in New York -- how long have we had progress reports? How long?

MR. HUGHES: Four years.

MR. MULGREW: Four years. And they've changed the weighting constantly in trying to get the peer groups right. But all of the principles -- you can sit there with them, they will tell you there is a disadvantage. And the schools that have been slated for closure all have high percentages of those students. So they think it's just a disadvantage to have that type of student, and that's just wrong. We shouldn't have a school system like that.

MR. GREENSTONE: I think Bob wanted to -- yeah.

MR. HUGHES: Yeah, I appreciate the kind of ambition of the strategy. And when you positing the question as choosing between two schools in a competition, one you close and one you open. I mean, New Visions is on record as supporting school closing in controversial times, and we stand by that.

I do think getting that decision right is very hard for the reasons we've talked about. The nuance of measuring the data. But also, the capacity of the organization making the decision. So you know, frequently school districts will look -- and just as a psychometrician puts the cut score politically because that's how many kids they need to have pass or fail, I think districts inevitably choose the number of schools they're going to open or close based on a political calculation of where that cut score should be. And a legitimate question about the capacity they have to undertake the types of reforms that will substantially change what happens in the building.

So you know, I would disaggregate those two questions, respectfully --

MR. NEAL: So if you look at these first --

MR. HUGHES: Just let me finish:? And say, yeah. We may get a more accurate sense of what's going on in the building, but then I think whether we want to use a single measure or multiple measures for those decisions about closing schools is an open one.

We need to close failing schools. But the decisions need to be really --

MR. NEAL: Use multiple -- obviously. But if you -- I calculated these for Chicago for -- they have like 400 elementary schools in Chicago. And if you looked at the percentile performance

indices for Chicago, you know there would be about 10 schools that would stand out on this graph where you would go, you know we need to go look at these closely and get these multiple measures.

And the interesting thing is that in some years if you take their VAM reports, there are a couple -- they've got these huge, enormous negative VAM scores that don't look bad according to PPI index. And so, I think it's important to create multiple statistical measures, to create multiple measures based on school inspections, et cetera. But have this commitment to relative performance so that the standard you're judged against is how well people who are willing to work with the same students are doing and not some target that comes out of a formula. Because those targets can be manipulated and those targets can be handicapped wrong so that ex post you don't work -- want certain types of students.

But if it's always a fair fight, then the people who are best with special ed kids are going to want to work with the special ed kids.

MR. GREENSTONE: Okay, I want to move the conversation a little bit. Let's -- taking this important idea. And now I'm going to cite a statistic which I'm not certain is true. But Adam Looney in the back of the room has told it to me many times. That in the next X -- I think X is 10 years? We have to hire 5 million new teachers. You guys probably know if that's right.

But let's just say it is, and let's say it's in the ballpark.

MR. MULGREW: Are you guaranteeing budget on that? (Laughter)

MR. GREENSTONE: I'm guaranteeing the budget that I control on that.

How does all of this fit into attracting, you know, from the broadest pool possible of teachers? And what is the way that we can use these assessment-based systems as we're continuing to try to learn how to operate them best in a way that will make sure we're getting the best and the brightest into the teaching profession?

MR. GORMAN: Well, first comment I would say is I think the biggest problem we have with the accountability for adults is we don't tell the truth to adults. It's funny, we'll tell the truth to kids that they're not performing but we don't go and tell folks this is what you're doing really well and this is what you're struggling with. And let's work on that development piece. It's just -- it's very difficult to tell the truth to adults, it seems.

And if we just started with that as step one, we could move a heck of a lot farther down

the pipeline without even doing some of these other things. Just from what we've got already. It's just difficult.

Related to that piece, that whole HR pipeline for school districts is really not one of the higher-functioning parts. If I think about the way we on-boarded at Charlotte would be an utter disincentive for someone to say, I got hired last week, I'm one week in, man have you got to come and join me here. That just wouldn't happen. Everything from the impersonal way we hire and on-board and things like that.

So I think one of the first pieces we've got to do is look at that whole HR pipeline for how do we recruit, how do we hire, how do we place, how do we professionally develop, how do we evaluate, how do we retain, how do we reward? And if we're going to tackle this issue, I think we've got to approach each one of those separate points along the way. And it's not just the one of how do we evaluate?

Because I use this -- I was on the panel with Randy before and I said, we're not going to fire our way to be a great district in Charlotte. That's not going to get us there. It's going to be almost more important for how we hire and coach everyone we've got along the way.

So I think it is this discrete piece of tackling every part of that cycle along the way and involving teachers in the development of these solutions for how we do that. The best systems we've come up with so far in Charlotte were developed by our principals and our teachers working together, far and away.

MR. HUGHES: Yeah, I would just echo that. I think that when we look at New York City, one of the things that we found is that teachers who leave who you want to keep leave because they don't feel the school has made an investment in their professional growth and development.

Everybody wants to be on a team, everybody wants to win. And we frequently don't focus on those managerial strategies that are going to keep good people there and weed people who shouldn't be teaching out. And so we need to do that.

New York did a study a couple years ago and it found that almost 67 percent of the principal's time was spent doing administrative activities that were unrelated to the building and were created by authorities outside of the school. Then when you add the actual operation of the school, it was

something like 15 percent for instructional time. So we somehow have to get our priorities right on this, otherwise we aren't going to make it.

MR. MULGREW: And I agree with both of your statements. I would take it even a level further. The teachers are spending more and more of their time out doing work that has nothing to do with instructional practice also. And that is the most important thing that they do.

So in terms of the HR pipeline, I am convinced personally that if we wouldn't -- if we weren't in the economic situation we are in right now we would see a complete fleeing of the teaching profession because of all of the politics that have been put around teachers right now.

I go into schools constantly and they're just like, why are we the enemies? That's not good for recruitment, you know? Why am I lazy, greedy, and hate my students, as the papers portray me? All I wanted to do is sit in a classroom all day with 25 to 30 children and help them learn. That is not going to bode well for recruitment in the future.

Having -- and Bob is 100 percent correct. On the exit interviews they all say the same things. I do not feel that there are people and there is a system designed to help me help students. And that is why I am leaving this profession. And until we tackle that, we're not going to tackle the bigger issue on education.

MR. GREENSTONE: Okay. I think we'll turn to questions from the floor. Derek, I think that there was -- there's -- this came across in a couple of the questions that came over. Maybe you could talk a little bit more about the ways in which critical thinking skills could be incorporated into your idea?

MR. NEAL: So, let me be clear. I've never developed a test. Okay? What I know from contract theory and economics is that the whole idea here is to design a system so that when the teachers act in their own self interest to maximize their own job security, their own paycheck, they do exactly what you want them to do in terms of best practice.

And so you're trying to reverse engineer a contest so that the best strategy is the best thing. And my grand vision is that if people knew and trusted some independent testing agency, that every year the exam was going to cover the whole curriculum. But from year to year, it was going to be very different in terms of the mix of essay questions or fill-in-the-blank or possibly, you know, requiring

you to synthesize and do case studies. You know, different type -- then the only way to say I'm going to prepare my kids effectively for this is to try to teach them in ways so that they can communicate their mastery in a lot of different settings. Which is what I think we want.

And one of the questions that I would have is, do we think there's the political will to have an independent testing agency that was not giving students what they expected but really kind of challenging -- and because there would be some students some years that would under-perform simply because they didn't kind of match with this year's way it was set up. And there would -- so there is a question about the political will to have a testing system that is less predictable, because I think that as much as everyone hates the predictability and the coaching, I think sometimes teachers take a little comfort in at least they know how to prepare for it if it's going to have high stakes against it.

And so I think that's another one of our initial comments were about political will. That's one of the things we have to gauge, is whether there is the will to really put people in this contest where the only way to prepare is just to try to know it well.

MR. HUGHES: But it strikes me in your comments -- that was helpful to get a better sense of it for me.

The working conditions we put teachers in right now need to fundamentally change if we're going to be serious about what -- holding them accountable for what kids know and are able to do. And I mean that in two ways. The way you described assessment I think is too frequently the reality. That teachers can predict some of the test but not all of the test. And there's no consensus on what should be happening in the classroom. So we typically don't have scope and sequence, we typically don't have curriculum in our schools. And I think starting to get clearer on what we want teachers to know and teach would be helpful.

Because I think the second point for me is, teaching has traditionally been defined as an individual profession. And I think you're starting to see in Charlotte, Mecklenburg and New York the best schools conceptualize teaching as teams of teachers working with groups of students for an extended period of time with effectiveness at the core of professional identity.

So getting to that place is important.

MR. NEAL: I agree on both of those, and everything I write works on the assumption that

those things have already been worked out and fixed. (Laughter)

And it was -- I wrote -- in the first paper I ever wrote on this I was asked to come down to the thing at Vanderbilt, the Center for Performance Incentives. And they wanted an economist at their first conference to write a paper. And I started with it has to be -- you have -- the first thing you have to do before you do anything else is you have to have a detailed curriculum and communicate clearly the importance weights on each part of it.

That there's no way you can design an incentive system to get people to do a certain thing unless you give them clear guidance on what they're supposed to do. And if you don't do that first, then -- and I also talked about the importance of teams and sharing information and coaching each other up. And everything I say assumes that those two things are fixed before we start.

What I'm suggesting doesn't really help much if those two things -- I assume that people were already doing that. So.

MR. GORMAN: So a new title for this panel is, Utopia Meets Dystopia. (Laughter)

MR. MULGREW: We had to push in legislation when we changed our valuation legislation to put a line in that teachers have to be supplied a curriculum. People thought I was nuts. I said, no, they don't have it. This is a problem.

And the fact that we had to put it into legislation -- and I'm glad we're able to get it there -- but just shows you how sometimes -- how sad it was, that that had to be put into legislation. And we have a lot of people breaking the law right now.

MR. NEAL: With all the stuff about state standards under NCOB, I just assume that every school had a big curriculum from the state. (Laughter)

MR. GREENSTONE: Okay.

MR. MULGREW: You want to go back to the federal policy issue again? (Laughter)

MR. GREENSTONE: Okay, I think we have time for one more question. And this question has been asked in various forums on a couple of these cards. And I think this question must be driven by the statistic that Eric Hanushek waves around that if we could just find a way to remove the worst 5 or 6 percent of teachers suddenly and replace them with a median teacher, 50th percentile teacher, we would -- our school systems would perform as well as Finland or whatever the top of the

rankings are.

And the question is, you know, what are the political constraints and how can we remove them to taking the most readily identifiable teachers who are not performing particularly well and bringing new teachers in --

MR. MULGREW: This would be for me, right?

MR. MULGREW: Two years ago in New York City, there were the famous rubber rooms. So I said, this is not acceptable, this doesn't work for anyone. This is ridiculous. So we went -- I went and I pushed and pushed and pushed so we could close them. So yes, we were the pushers to get them closed.

It wasn't just closed. I said, when I looked at all of the issues inside of it, I said, this is absurd. My responsibility is to give people -- to make sure due process is being followed.

The law was clear, but it was the exceptions in the law that were being followed all of the time. So the changes that we made, I'm -- you know, New York City, you know people don't speak about it now. But once some allegations or charges have been made to -- for a teacher, the average case now is 38 days.

When -- with insiders signing that agreement, our backload -- we remove the entire backload of New York City that was there for years and took up so much of this educational debate. We removed it in four months. And 188 teachers left as a result of 30-28 charges -- that's a removal of license -- in New York City. Yet you never heard anything about that.

So making the fair and fast process -- and now right now I am in -- I am trying to -- we don’t need the arbitrators we have anymore because all the cases are so quick. And the cases that go full 30-28, the average case is 92 days and it can't go over 105. So there is a way to do that, but we're not seeing the results of the piece that you said, that this will now make us Finland.

You know, Finland doesn't have standardized tests. Finland has a longer school day, but the children spend less time in school than the children in the United States. Their longer school day is based on the premise that teachers are working together and planning collaboratively on instruction. You know, you never hear that in the debate. Let's make a longer school day, like these countries have. But the children aren't in the school. They have less seat time than our children. It's about doing real

planning around what happens inside that classroom.

So I think that's what will get us to Finland. Not the other piece. And I will be more than happy to show anyone all the statistics that we have. We have proven that you can do this fast and fair, just -- it took the political will. And it was not an easy political will to get through, because there were people on the other side who liked that idea of having that issue out there and using it politically in the media. And it was really a push to say, this just does not work for the kids. It doesn't work for schools, it doesn't work for the running of a school system, so we need to fix this. And we were able to happily get to that agreement.

MR. GREENSTONE: Did you want to talk about this?

MR. GORMAN: Yeah, I would just quickly say. You know, to me it comes back to something I alluded to before. If we don't tell the truth to adults, we could be lying to kids.

SPEAKER: Right.

MR. GORMAN: We could be telling them that the teachers are doing what they need to, et cetera. But we need to coach and help them along the way as well. But if we start off early being truthful to people and look at how many schools do we look at -- and if you have a 4-point scale, 98 percent of the ratings given were a 4. What kind of diagnostic sitting down with the staff and giving them coaching and development? If we just started off with telling the truth. And I think that piece can start with the leadership of the district and working with the principle and some other pieces of just telling the truth.

I know it sounds -- so that's a common thing. But we don't.

MR. NEAL: And a commitment not to have scales, but to have relative performance rankings puts discipline and forces people to tell the truth.

If no one remembers anything else I said today, you always want these performance measures to be contest-based because then you can't systematically lie for everybody. We need to take Lake Woebegone effects out of accountability in the United States. We have to have systems where we commit upfront that everyone can't be exceptional, or we will continue to have fudging and distortion.

SPEAKER: Correct.

MR. GORMAN: I couldn't agree more. Because the way we started that telling the truth

piece was, I sat with all the area superintendents and I said, rank order and effectiveness all your principals. And they gave me the rank order list. Then I had HR pull the evaluations of those individuals. And I said, so how come this is not Lake Woebegone? Because rank order person number 1 and person number 25 have been evaluated exactly the same. What are we doing to coach those folks? We're not telling them the truth and giving them the support and help they need.

MR. GREENSTONE: Okay. You know, I'm just -- so as Roger said, you know, we chose to do another event on this. I think -- we think that there's no more important issue facing the country. The stagnation of test scores, the decline of living standards for many Americans -- and this has got to be part of the solution, if not the primary part.

And I'm just inspired to be onstage with you four, and the other people who were here earlier today. And you know, this is just fantastic. And we will be back with more research in a couple years. But for now, if you could join me in thanking the panelists.

(Applause)