

**THE BROOKINGS INSTITUTION
BROWN CENTER ON EDUCATION POLICY
BPEP CONFERENCE ON NATIONAL STANDARDS
May 15-16, 2000**

***Brookings Papers on Education Policy 2001*
Diane Ravitch, Editor**

**The Brookings Institution
1775 Massachusetts Avenue, N.W.
Washington, D.C. 20036-2188**

**TRANSCRIPT OF CONFERENCE
GENERAL DISCUSSION**

CONTENTS

CONFERENCE PAPER	TRANSCRIPT PAGES
Why Business Backs Education Standards	2
Authors: Milton Goldberg and Susan Traiman	
Comments: Alex Molner <i>and</i> John Stevens	
How Good Are State Standards?	11
Authors: Chester Finn and Marci Kanstoroom	
Comments: Richard Rothstein <i>and</i> Bill Honig	
Searching for Indirect Evidence for the Effects of State-Wide Reform	22
Authors: David Grissmer and Ann Flanagan	
Comments: Uri Traisman <i>and</i> Rob Meyer	
The Controversy Over the National Standards	31
Set by the National Assessment Governing Board	
Author: Mark Reckase	
Comments: Michael Feuer <i>and</i> Ed Haertel	
The Role of End-of-Course Exams and	38
Minimum Competency Exams in Standards-Based Reforms	
Authors: John Bishop, Ferran Mane, Michael Bishop, and Joan Moriarty	
Comments: Richard Murnane <i>and</i> Larry Steinberg	
Incentives and Equity Under Standards-Based Reform	43
Authors: Robert Costrell and Julian Betts	
Comments: Meredith Phillips <i>and</i> Herbert Walberg	
Academic Standards and the Social Structure of High-School Achievement...	54
in Shaker Heights, Ohio	
Author: Ronald Ferguson	
Comments: Jens Ludwig <i>and</i> Wilbur Rich	

Monday, May 15, 2000

Why Business Backs Education Standards

Authors: Milton Goldberg and Susan Traiman

Comments: Alex Molner *and* John Stevens

GENERAL DISCUSSION

MR. ROTHSTEIN: I wanted to comment briefly on this last exchange that Susan Traiman -- about the relative growth of high-skilled and low-skilled jobs. It's certainly true that the high-skilled jobs have the fastest rate of growth. But of course it's also the case that if we take a small base and apply a large percentage to it, we get a smaller number. But if you take a large base and apply a small percentage to it -- And while nobody has really figured this out, I don't think there's much evidence that overall, job vacancies in the future, when you combine both vacancies in new jobs and the placement vacancies from existing-type jobs, are going to be unbalanced, more highly-skilled by a significant -- probably a little bit -- but by a significant extent than in the past. As you pointed out, there are still going to be a lot of low-skilled jobs in the economy. The more interesting point, it seems to me that you made, was that there's something very deeply undemocratic about pre-determining in advance who gets the opportunity to compete for those better jobs, given that there are going to be large numbers of jobs that are not as desirable. And it seems to me that if education reform is predicated on making that competition more democratic, more meritocratic, having a more meritocratic society in which the rewards and the lack of rewards are based on equal educational opportunities, then we have a lot of thinking we need to do about other aspects of the organization of the society. Right now, we use education as a sorting mechanism, as a way of pre-determining who is going to get the better jobs and who is going to get the worse jobs. If everybody has more relatively equal educational qualifications, we have got to have some other way, both of explaining to people why they don't succeed in that competition, and of making the consequences of not succeeding less severe. For example, in Japan, which is often thrown up as an example in educational reform debates, one thing about the Japanese economy, is that you can support a middle class family in a fast food restaurant, a job at a fast food restaurant. If we want to have an educational system in which workers in fast food restaurants have educations which are more relatively equal to those with high-skilled jobs than is presently the case, then we also have to deal with the fact that in this society, getting a fast food job is not an acceptable alternative, because the compensation for it is so minimal, compared to the other jobs. These are not simply educational issues. We cannot create a meritocratic educational system while having such a vastly unequal economic system.

MS. RAVITCH: Just one of you can take on -- if you want to respond, or just let it stand as a comment.

MR. GOLDBERG: Well, I think it's a comment that's certainly worth discussing. And the problem is that I don't see an alternative of not -- with making it -- not making it possible for every youngster to aspire to a first-rate education. I think you've got to begin with that premise before you tackle the others.

MS. RAVITCH: John Bishop?

MR. J. BISHOP: I am amazed that I hear this argument. Seventy percent of the growth of jobs in the last twenty years has been professional managerial jobs. That seems to -- it's not a -- it is a fact -- professional managerial jobs, as a rule, grow at twice the rate of service jobs. The presumed McDonald's jobs that were -- that are the big growth jobs in the U.S. economy, operative jobs and labor jobs have been either flat or declining. So basically, the low-skilled jobs in the economy are a falling share of total jobs, and the high-skilled jobs are the growing share, and account for most of job growth, point number one. Point number two is that you complain about the wage that low-skilled jobs hold. It's because of our educational system that they are paid so poorly. Wages are determined in demand and supply markets. And it's because we have too many low-skilled people relative to the demand for low-skilled jobs, the supply of jobs for these people. Their wage is low for that reason. That's the major reason, in an economy where we don't have a minimum wage that forces the wage for low-skilled jobs to be a lot higher. And if we market labor market, like the one we have, it's -- that is, the educational system, and the fact that we have a large number of unskilled people, that's the -- fundamentally, the reason why those wages are so low, and of course, because we let -- we have additional people coming into the country with low skills, in addition. Another argument that Alex made that -- I think the argument that the business community needs better education in order to be profitable was never a correct argument, although I can see why PR people might want to make that argument to the business community, because you need to build their support. Business people figure out how to be profitable under any set of circumstances that exist in the U.S. And if the education of our people is poor, then we'll figure out ways around that, by moving plants abroad, by importing skilled labor, et cetera, and they are doing that right now. So this is not an issue about increasing the profit to American business. The issue is the interest of unskilled and lower skilled people and their children. That's what at stake, because if we continue with the low-quality educational system, they themselves will find themselves competing with lots of other unskilled people, and their wages will be lower, and their own children will have fewer opportunities for improvement in their occupations in the future. The argument that I have seen in papers, and referred to here that, "Gee, our economy is so productive and so successful, that it must be wrong that we have a problem in our education system." Two points. One is, is the part of the economy that's been so productive and so successful has been the college graduates and the people with high levels of skill. They have got most of this extra money that the economy is producing. The relative way it -- there was a calculation done a while ago of who got the increase in the GDP in about a 10-year period. And almost all of it went to the people in the top 20 percent of the wage distribution. So consequently, there has been an improved economy, and it's to the benefit of the high-skilled people. And we have had a widening of the distribution of wages, partly as a result of the market wanting more and more of those high-skilled people. But the other point is that the education system is partly responsible for recent improvements. There has been a big improvement in the educational system in the last two decades. From the bottom, in 1980, grade level -- the math and science scores of kids coming out of high school have gone up on more than one grade level equivalent. Now, in the economy, a grade level -- people with one additional year of schooling, get about 15 percent extra. Much of that higher wage rate is partly also working more. Some of that is due to other correlated factors. Let's say the causal effect is 10 percent. So essentially, the people coming into the labor market in the early nineties, versus the people coming into the labor market around 1980, were about 10 percent more productive. And that's part of the reason for the improved productivity of the economy as a whole. Another way education has contributed to the improvement of the GDP growth in the economy is there's been a big increase in the number

of people graduating from college. The ratio of degrees awarded is now like, 35 percent. So consequently, there's been a very, very large response on the part of students going to and staying in college and completing college. So the educational sector is helping improve and helping growth in the economy. But there is a part of the educational sector, that which serves the kids in poorer communities, in high school and middle school, that is doing a very poor job, and that's the part of the sector that the standards-based reform is designed to improve.

MS. RAVITCH: Maris Vinovskis?

MR. VINOVSKIS: Let me take a little different perspective on what Alex has been trying to say, because I think there's a real danger that we are exaggerating, and that it's going to come back to haunt us, the economic productivity of education. And this isn't something that's new. One of the things that educators in our history have done well is that whenever we face a problem, we come up with rhetoric and ideas to solve the idea of economic productivity. Let me give you an example, and you tell what I'm concerned about. American is actually the nation that leads in the idea of economic productivity in education. Adam Smith and others really didn't push it. It was Horace Mann. He didn't start out that way. Horace Mann was interested in character education. It didn't suck. His board was almost abolished. So what did he do? He invented, literally, the idea of economic productivity. He falsified the data to make it look good, and it did look good, and we got on this kick of economic productivity of education, which I'm not against. I'm all for standards. But the difficulty is that to sell that reform, he did it in the name of economic productivity. Having now sort of gone through repeats of this through our history, I've been looking at the materials in the 1980's, and reading some of the things that some of you have been writing about economic productivity and education. And these are vast exaggerations. You look at the southern governors. Not -- they really were pushing because they believed that. Now, I mean, it's a relationship, but it's much weaker. Human capital analysis is much more subtle about these things, and I think Milt and Susan point to that. The difficulty is, if so much of our reforms have been sold in the name of productivity, and oversold, what's going to happen when Richard Rothstein and others -- and they have been doing this -- are beginning to say, "Well, wait a minute. It's more complex. It's not that there's no relationship, it's more complex. What happens if we succeed, and the American public becomes littered in economics and begins to see this?" Can we sell -- and I hope we can -- standards-based reform on the idea of equity, on the idea of equal opportunities for people? So my fear is that the business community has a fairly simplistic view in its publications, which is perpetuated all too often, and that we're not going to see this broader concept, which I think Milt and Susan are now starting to do. But we have to do that, because we need these reforms, but they're not going to be the simple reforms that we talked about in the eighties and nineties.

MS. RAVITCH: John?

MR. STEVENS: I would just like to say that, in Texas, we have been very concerned about equity. Part of it is demographics, as we've looked at the future of the state. The demographic makeup of Texas is dramatically changing, and we recognize that. Business has been at the forefront of saying, "We must educate students who have historically done poorly in our schools. They must be able to participate. Not just because we need them, but because it's the right thing to do." We have a standards-based approach, as you know. Schools are accountable for the

performance of each student, separately. And that has made a dramatic difference in closing the historic performance gaps that existed out there. White students and economically-advantaged students have shown progress, African-American, Hispanic, and economically-disadvantaged students has been much more rapid. And part of, I think, what we need to look at, is the value of this kind of an approach to closing those performance gaps. Not just because of the demographics, but because we do want a democratic society where people have an opportunity to participate. I don't see that the other approaches are going to get us there. There are always some strange alliances in business, and I was just amazed. Those who are really worried about the disadvantaged in our society and those who are already advantaged, have once again joined forces. They're happy with sorting, for some reason, and that's really not what this ought to be about.

MS. RAVITCH: I --

MR. MOLNAR: I'd like to respond to that.

MS. RAVITCH: Alex?

MR. MOLNAR: I want to come back to something that John Bishop said. I'd like to go over something that's just been said. I think it's a circular discussion, to talk about that low-income workers are competing with each other and keeping each other's wages down. And if only if we could propel more low-income workers into the more rarified atmosphere of high-skilled workers, that low-income workers would have more weight in the competition for low-skilled jobs, and drive wages up. By my reading of the hydraulics of that analysis, that would mean a general reduction in the wages of high-skilled workers. So therefore, equity would be produced by driving down the wages of high-skilled workers, and bringing the low-skilled workers' wages up, which isn't quite, I don't think, what most folks had in mind when they think about this equation. Well, I think it's one of the problems, this entire discussion is by and large dominated by economists, some of which I find appealing and some I find appalling, but all of whom I found rather narrow within their focus. Let me come back to the issue of the standard of what's the kind of world in which the standards are embedded, and do we wish a world in which certain occupations are stripped of dignity? I take it as a given that we'll always need folks that we describe now as occupying low-skilled occupations. Does that mean their children must? Of course not. Human talent isn't distributed by social class. It would be the worst thing in the world that we could do, to choke off opportunities for every possible sector of our society to contribute to the fullest of their abilities. The question is whether the adults who rear those children lead lives of desperation which effectively foreclose the possibilities of their children, regardless of standards, or schools, or anything else. So the question of schools not performing in areas is an important question. But it is not more important than the question of the wretched housing, inadequate health care, and the nutritional deficiencies that poor people in this country, and working poor people, live with every day. And we need to make that a much more part of the standards discussion. We have to make the standards discussion, as has been suggested, much more complex. This is really -- the discussion we're having so far, if I may say so, is a nineteenth century discussion. You can almost hear the machine clanking. You can almost see the arrows being drawn to the rectangles on the boxes. Now as I take it, the new knowledge industry is characterized by much more chaotic environments, much more chaotic systems, much less sort of

straight line, linear thinking than we've heard so far, today. If you want to do something effective about standards, why not the hell get rid of secondary schools as they're currently constituted? I don't think they can be redeemed. Not by standards, not by anything that's been mentioned here. But that would require a real transformation, not only of occupations, but the way in which we integrate adolescence into adulthood in this society. That needs to be talked about.

MS. RAVITCH: Chester, you've had your placard up a while. Is it still up, or did you take it down?

MR. FINN: Well, I've taken it down, but I'm going to say something anyway. I do think that this was a good paper to open with, because it seems to me it's framed very nicely. The -- what I believe is the philosophical dilemma underlying the whole conference, which I don't think is going to be resolved by economists or by data, which is the philosophical choice between the world view, which says high standards are very important for all kids, even though it's going to be painful to get there, and a world view which says that someone's got to take out the garbage. And -- the "But someone's got to take out the garbage" argument, which Alex made earlier, which has been made frequently by Gerry Bracey and by a number of others, usually from the political left, brings with it all these -- and was rebutted by Susan, I think, with her kind of upward mobility argument, how at least your children should be able to escape from the garbage truck by a good education and a chance to go to law school, and I think this is a very important framing argument for the whole discussion. I've always been sort of perplexed by the taking out the garbage argument, as -- especially why it comes from the left, because it always strikes me as so fatalistic, deterministic, ultimately, probably even racist, given the way it works out, at least in this society today. But it's wrong to frame it just in job terms. It's clear in Milt's and Susan's paper, though it's not their emphasis, that the education they're arguing for isn't just for job and economic reasons. It isn't just to get a high-tech job. It also has to do with socialization, culturization, civic engagement, all of the other things for which we presumably send kids to school, educate them at home, want them to get educated. There too, it seems to me, there is an argument for high standards for all kids. And I don't know what the civic counterpart of taking out the garbage is, but I suspect that it's not voting, not reading the newspaper, not engaged in the culture, not reading books, not being a very good neighbor or citizen, or probably parent. And it just seems to me that there must be a kind of non-utilitarian counterpart to the someone must take out the garbage argument. I think it's worth surfacing that and seeing whether we want to embrace it as a maxim for American society.

MS. RAVITCH: Thank you. I'm going to call on Robert Costrell next. I just want to ask that sort of as a general blanket rule -- and it's a request, not a command -- that in the future, no one refer to their social class origins or their -- you know, "my mother, my father."

(Laughter.)

MS. RAVITCH: I don't want to have to do that myself, so I'll ask if you can please refrain from doing so.

MR. COSTRELL: First of all, with regard to how much garbage we make, I think it's also worth remembering they make a lot more in a productive, highly-educated economy than in an economy

where it is not -- there is not a high level of education and skills. There's complementarities across the skilled distribution. But that's not the main point I wanted to make here. With regard to Maris's point, I did have some sympathy with the concern about the exaggeration of certain aspects of economic benefits of improving education. And I think it was some exaggeration with regard to the argument about why we were trailing Japan in the early eighties, which clearly did not pan out. But the equity arguments are economic. And I think you mistakenly implied, or at least I heard -- seemed to imply -- that they were different, as John argued quite, I think, accurately. The equity concerns are at the face, economic, the demand for people of different skill levels. And the technical bias of -- the bias of technical progress that we've observed over the last 20 years, which has led to the widening that John spoke of. But I want to -- my main question is actually to the representatives of the business community here. Because for all of these discussions about how the economy works, and how the economy doesn't work, I think the acid test of Alex's hypothesis that you know, it isn't going to make a difference, economically speaking, is going to be this. Will, or do businesses make state location decisions on the basis of the strength of state standards? You've talked quite a bit about international comparisons, and I think you might want to think about shifting from that to comparisons amongst the states. I mean, with regard to the international -- and by the way, I can't help throwing out one line that I read from -- what was it -- Berliner & Biddle, the manufacturer crisis. It was actually commendably candid, in their statement that, "Well, we don't really need to raise our math and computer skills because there's a lot Pakistanis out there and Indians out there who we can import." Of course, I didn't yet hear them weight in on the H-1(b) Visas, but maybe they will. But I think that the issue has to shift back to the states, the location decisions of the states. I can tell you, in one of the hats that I wear -- I know I'm not supposed to give my -- this is not my class background, but my current professional situation -- I am now on leave from the University of Massachusetts, and working with state house, the Commonwealth of Massachusetts, and this is incredibly important. Incredibly important, how these decisions are being made on a variety of dimensions, tax policy, any other policy. This is what will carry freight. This would be the acid test. If you're engaged in the backlash issue, this is what's going to be the -- and in fact, it's the way it should be, because the federal system that we have of states trying different policies, of seeing how they pan out, is -- the way it works is because of the mobility of decisions that are made. So that's my question to you.

MS. RAVITCH: Answer, and I'm going to try to get to the other three people who have their placards up.

MS. TRAIMAN: This is starting to happen. We know of states where, in surveys of manufacturing companies, companies have said, "We cannot grow in this state, because we don't have a sufficient number of qualified people to hire coming out of high school." We also know that companies make their location decisions on a variety of issues: the regulatory environment in the state, the tax policy in the state, and that as states vie for businesses to locate, they put out all sorts of goodies that companies weigh as they decide where to locate. But we have increasing evidence that one of the factors that companies are considering is the availability of skilled people for jobs and they're looking both at high school graduates, but also at the infrastructure for post-secondary education.

MS. RAVITCH: John, say it in 60 seconds.

MR. STEVENS: I would say that decisions as to where to locate are driven by a lot of factors. Part of the question is who participates when they decide by tax policy, transportation, a lot of other factors where they're going to locate. But another very important factor is, "If I have to recruit people from outside, executives, engineers, and so forth, what is the quality of life in the community that would be able to attract those people?" And the kind of schools that are there, not just for our kids, but in general, are very important in being able to bring in, if you have to do that, the people that you need to operate. So there's no getting around the fact that education has an enormous impact. It's not everything, but it's one of the major factors. Not just for the equity standards. Standards is not results. That's what matters, you know? The results is the bottom line.

MS. RAVITCH: Kevin Kosar

MR. KOSAR: A question for Alex. If for the most part, businesses don't need more learned workers, why then are they pushing for raising achievement? Are they just misunderstanding their true interests, or is it some sort of grand ruse, or what?

MR. MOLNAR: Oh, I'm willing to stipulate that businesses need more educated workers in a variety of different occupations. It's a question of the relative degree to which this discussion, with regard to the performance of schools and the development of standards drives that equation. There's plenty of evidence of businesses in this country laying off highly skilled workers, highly trained workers, and relocating to Bangalor, India, to Mexico, and so on, where they can simply get cheaper labor. And they do that. So what I want to suggest is, is that I think that many business people believe their own rhetoric with regard to standards and wanting a highly trained, educated work force, I'm simply saying that the discussion, when it is constrained to an economic discussion, in which it's driven by business interests, is necessarily an incomplete discussion. It's narrow, and it focuses on special interest. If I understood John's response, at least in part, the role of education with regard to the placement decisions of corporations and industries is a quality of life decision for corporate executives. In other words, the argument wasn't that the schools in the area would necessarily provide this, that, and the other job, but that executives would be attracted to areas because their children could find decent schools to be educated in. It became a quality of life consideration, didn't it? I would suggest again, there's a lot of value in thinking about standards, and there's, "Who will be against highly educated, well-skilled people?" We're not -- I'm in favor of that. I've never been against that, never have written against that. The point again, has to do with the dignity assigned to folks. This equity argument is a tricky one. It keeps getting redefined as we go around the table here. It's slippery. We need to address that some more, and we need to focus maybe a little bit on some concrete, tangible things that corporations additionally can do to increase the standards in schools. Let me say that one thing they could do tomorrow is stop advertising kids in schools. That would be a dramatic thing that they could do that would increase standards immediately.

MS. RAVITCH: Okay, we're going to --

MR. GOLDBERG: May I just add one -- I just want to say one thing about the children of the executives going to school. It's the children of the employees of the company. We're not talking

about the executives here. I mean, Susan's organization, my organization -- The fact of the matter is, there are 34 million employees of the companies that are members of our groups. Thirty-four million employees, that's who we're talking about when they talk about who goes to the schools in these communities.

MS. RAVITCH: We're going to do this. I'm going to ask John Hoven to make his comment or question, and then I think -- is it Sally Kilgore that has her -- and then Julian Betts, and we'll get each of your comments or questions, and then we'll either get a response or end it. But we've got to move on. John?

MR. HOVEN: Comment and a question. First, a comment. I think it's important to keep in mind that a concern that standards had to address is -- applies to kids across the entire achievement spectrum, gifted, average, and below-achievers. We know that our top 10 percent of students, the mathematics are scoring at the level of the median student in Singapore. The example that Susan gave from Montgomery County public schools, where in one school, 99 percent of the African-Americans failed ninth grade algebra test. Those were not the low achievers. The low achieving African Americans aren't taking algebra in ninth grade. Those are the average students. County wide, two-thirds of all students failed first semester algebra. So there's a need for higher standards at all levels, gifted, average, and low achievers. And the question. The -- my school district has used internally-designed standards as a device to make itself look good. So there's a need for externally designed standards. And one way of doing that that is important among businesses is benchmarking. And I don't hear this talking about urban schools who benchmark themselves against the best educational school districts or schools they can find outside of their districts. Is that a gap? Is that something that's going on I haven't heard of? There's something about to be done.

MS. RAVITCH: Well, I'm going to ask you just to hold the answers until we get all the questions out. Sally?

MS. KILGORE: I guess mine is -- what should I say -- a comment/assertion. What's to say that the assumption that both sides of the argument today have is a nineteenth century view of the economy? And that really, we are in a -- that nineteenth century assumption is that there's a certain fixed distribution of opportunities that, you know, may change to some degree, but in other words, there's a constraint, because people are going to an employer. And what's to say that the -- what you're in now, and have had in the future, is one that is created by individuals, that is to say, outside of large corporate firms, and you're able to create jobs, but also because the lack of constraint on trade that the Internet produces. First of all, by trades, electronic transmissions, of things that were locally circumscribed, means that the mere notion of the economy that we have now as a fixed distribution of opportunities is gone, and to the degree that we're going to act on it, we have an unlimited set of opportunities, which -- an unlimited set of meritocratically achieving students could go to.

MS. RAVITCH: Julian Betts?

MR. BETTS: I'd like to make a couple of comments about Alex Molnar's discussion. First of all, I hope you really have time during today and tomorrow to talk about one of his points, which is

that if we have teachers teaching to the tests, and the tests aren't the right sorts of assessment, they're simple multiple choice, rather shallow types of assessment, we will end up defeating creativity. And I hope that we discuss that at some length. The main thing I wanted to talk about is his contention that this is just a political education bandwagon, that we really are overstating the increasing demand for skilled workers, that really we've got creeping credentialism. And that just is completely opposite to all the facts we have at hand. If this were true, we would see people with bachelor's degrees struggling to find first jobs, taking jobs for which they were over-qualified, washing dishes and so on, and that's not what is happening. Wages for bachelor's graduates are not falling. In fact, since the late 1970's, we've seen roughly it doubling in -- the wage premium between college graduates and high school graduates. At the same time, we note the supply of skilled workers has been expanding. The only explanation for what's been going on is a throw at the 1970's, 1980's, and 1990's, there's been a very clear and steady trend in the demand for labor, to look towards more highly skilled workers, and you can't ignore that.

How Good Are State Standards?

Authors: Chester Finn and Marci Kanstoroom

Comments: Richard Rothstein *and* Bill Honig

GENERAL DISCUSSION

MR. MEYER: Thanks. I enjoyed the presentation, conversations. I wanted to address a couple of issues that came up, and I think we're going to come back to this issue of whether there should be a binary standard, or multiple cutpoints. I think that's an important one. But the thing I wanted to stress is the difference between evaluating students and evaluating schools. And I think that's an area where we need to work on integrating those two things. And I think Checker points towards the promise of using value-added models. To some degree, we need to be more successful at integrating value-added analysis and value-added models, so that they seem to link up and be talking about the same thing as individual performance. And sometimes, they're done in such a way that they're so technical, they seem removed and abstracted from the context of individual standards, and I think that's an important area to work on. And I gave a paper at ARA a couple of weeks ago on some ideas of how you might do that. I think that's an important thing to think about. Bill mentioned a point that you'd like to evaluate schools on the basis of how many kids passed the eighth grade -- the algebra test by the eighth grade, and I wanted to maybe take issue with that, and suggest that that's -- there's some problems with that. And I think, again, it comes to the issue of an attempt to kind of integrate a concern with how much schools are doing with a standards concern here, the standard being binary cutpoint. And I think there's a big problem with that. I think it's important to sometimes label points on the ruler, points that we're concerned about, but I don't think that that necessarily then constrains us to say we need to do our measurement, throwing away the other parts of the ruler. What happens if you have something like this, and many states are going to try to measure how schools are doing because they have this binary standard by looking at percent proficient, and so on, and my work in some of the urban school systems seems to indicate that when you have a particular cutpoint, it focuses resources on students around the cutpoint, some a little bit above and some a little bit below. And paradoxically, it -- these standards, you know, we've talked about them -- are really motivated, in part, to help the lower-achieving students, but you'll see some of the schools, not necessarily all, neglecting the students at the bottom, because they don't feel that they're quite ready to pass that eighth grade standard, that algebra standard, by eighth grade. So it's in the interest, the self-interest of the school, in other words, the society provides an incentive for them to take resources away from those kids, and put them to students who are just below the cutpoint or just above, because that's the rational way to maximize performance on that. And I think the problem isn't necessarily that it's wrong to label that point on the ruler -- although we'll get back to that, and Bob Costrell has written some interesting things on that -- but the problem is that that wouldn't, isn't necessarily the best way to measure how a school is doing. If a school -- maybe it had lots of its students -- took in students who were at the thirtieth percentile in sixth grade, and was able to raise kids to the sixtieth percentile in two years -- say the cutpoint is at the seventieth percentile -- I think you would want to document that success, and in fact, in documenting that, you'd want to find out why are they doing such a good job? They would be totally missed by that single cutpoint. In other words, that's not a very good way to identify how the schools are doing. So I think in general, looking at cutpoints, proficiency thresholds, is a poor way to identify how schools are doing. It focuses on just a few people. And the key here is to try to bring all these

things together, and are there ways of identifying how schools are doing at generating progress, and at the same time be sensitive to individual standards. I think that's an important thing, and maybe we can revisit that throughout today and tomorrow.

MS. RAVITCH: I think that, especially when we get to John Bishop's paper, and tomorrow, the Costrell and Betts paper, and Ron Ferguson's, we will continue to go into this -- these issues in greater depth. Ed Haertel?

MR. HAERTEL: Thanks, Diane. Just a couple of comments on several things that have been said. First, there are all kinds of pressures in the standards development process toward inclusiveness. The easiest way to get political consensus is to put more things in. In the textbook -- committee to select textbooks, they love to see if their particular content is there. They don't worry about the fact that there's other stuff that they don't -- that they're not interested in. Conversely, all of the market pressures and tests are toward going to the intersection, rather than union of different curricula. If I'm a publisher, and I want to make my market as large as possible, I'll try not to include anything in my test that any of my potential consumers aren't already covering. So the tests tend to get narrower, and narrower, and the standards tend to get broader and broader. If standards -- creating and celebrating and applauding standards is, in fact, a rhetorical activity, then we expect these documents to be very grandiose, and for everyone to be outstandarding their neighbors. I'm pleased that the discussants commented on the difficulty with skipping over the nature of the tests themselves, and going directly to consequences in the paper, because I think that the standards are -- the tests are the essence, the lynch pin of this activity, if it's going to work. Bill commented that all the tests correlate very highly, sort of implying that it doesn't matter which test you use, you'll get about the same rank ordering. That, unfortunately, is true. I've been working with the State of California, and I know that looking at school level correlations, which are not at all the same thing as individual correlation, these are ecological correlations using scores for school buildings, the correlation between mean parent education level and the performance index is above .9, .94. We can account for 90 percent of the variance in how well schools do, just by knowing the background of the kids that they serve. Part of the remaining 10 percent is measured in there. There's very, very little left over for any of these to account for. The hope is that scores -- not actual levels of scores -- will be less highly correlated with SES, but gains, or changes from one year to the next will tell us more about which schools are doing a better or worse job. But we know that changes in scores on one test are not going to correlate very well with changes in scores on another. The paper Bob Linn just published in education research earlier this year points out that gains of one high-stakes test don't even generalize to tests that look like they're measuring about the same thing in the same time area, let alone generalizing how well people can use these skills. So the precise nature of the criteria is very important, because it's at -- the precise criteria that's going to go up, and not all the other things that we'd like to think that criteria represents. Thanks.

MS. RAVITCH: Thank you. David Grissmer?

MR. GRISSMER: Yes, a couple of comments. I think it's important to keep in mind that, you know, we're not necessarily after high standards. It's high achievement that we're after. And standards are only a means, as part of a system, to get higher achievement, and a fairly complex system for getting higher achievement. So that I think it's really less important where standards

are set initially, although I think they could be set too high and too low. But I think in the long run, what standards do in assessment, is they provide the data that allows us to say what works and what doesn't work. And in the long term, that's what we need. I mean, we waste billions of dollars, because we can't answer those simple questions of why something works or doesn't work in the classroom, or this kind of kid, this teaching technique, this type of teacher. And the only way to ever get at those questions is to have the data available that the standards movement is producing, as well as assessments. And once we get that kind of data, the real purpose of it is to allow teachers and principals and researchers to use the data to improve the system. And in the long term, that's what, I think, this is all about. It's improving achievement, and standards are sort of the way to get from here to there. And so my sense is that the argument over high standards versus low standards isn't really a question at all. You want to be able to get the standards into the system that teachers teach to, that they have some respect, but basically it's to provide a measurement that you can begin to use, and teachers use. And I've seen it happen in the couple of states we've studied. Teachers look at scores from other schools that have the same SES students, and they're getting better scores over there. What are they doing? So that the reason we're going through this process is to eventually sort out what works and doesn't work. And the improvement in the system, a good deal of it will sort of result from that process.

MS. RAVITCH: Checker?

MR. FINN: Yes, I'd like to go back to a point I think was raised this morning, but we all know 75 percent of a variance in scores has nothing to do with schools, it has to do with families. And part of that is genetic, and part of it is what families do, in terms of providing resources, providing mentoring for their kids. So setting standards for schools leaves out the most important part of the equation that can lift test scores, the family. And so unless we can get families into a position in the country, and that -- the family variables that make a difference are income and education. So giving this generation of parents more education seems to pass back to the next generation. Income is a key part of the family part of this. Age of children, age of parent at child's birth, teen pregnancy. You know, we can get a lot of higher achievement out of family measures, than possibly out of school measures. So I think we really have to keep in perspective here how much schools can do and how much we can even expect out of this whole standards process.

MR. GRISSMER: Everybody carries in their head, and usually does not articulate what I've come to call a theory of change, some notion in your head about what is likely to cause behavior to alter by students, teachers, parents, what have you. I want to take mild issue with what I thought was the implicit theory of change you were associating with standards, which is the theory of change that would say that, faced with information, educators will spontaneously want to do better, and will change what they do in order to produce a better result. In other words, decoupling change from the accountability and consequences paradigm, and appearing to suggest -- and tell me if I misheard you -- that the main point of standards is to provide information on the basis of which change will somehow occur, even if there are not stakes attached to it?

MR. FINN: I think teachers have an awful lot -- you know, I don't think we're into a punitive system where you have to sort of, you know, force teachers to do what some don't want to do. I think teachers have a great deal of professional pride in their job. I don't think they've had the information they need to determine -- from researchers, in particular -- what works and what

doesn't work. I mean, you know, it's a zoo out there, as far as if you're a principal or a superintendent, or a teacher, and you want to say, "Well, gee, how could I improve what I do? Let's go to the research," you know, nobody home for the last hundred years. So that the sense here is that unless -- I think there's less -- that the theory of change here depends a lot more on professional pride, in giving people who are professionals the tools, the working environment, and the information they need to make the decisions, rather than somehow sort of a punitive model.

MS. RAVITCH: Larry Steinberg?

MR. STEINBERG: I want to make an observation and then I'll throw a question in. Observation goes to something that Bill mentioned in his remarks. I don't want to lose sight of this, because I think it's very important. It seems to be that one way to attract a dilemma that Richard pointed to, that is to say that the bottom of the distribution -- you're either going to have a lot of individuals who aren't passing the test, or you're going to have to set the standards so low that it's meaningless to have the test in the first place. One way to deal with that is to get out of the mode that schools have typically functioned in, in which we hold time constant and vary content, and instead, allow time to vary and hold the standards at the same level. So it's to say, if you could say by the time you finished, rather than by eighth grade, or ninth grade, or tenth grade, or whatever, you need to do X, Y, and Z, that would accommodate the fact that some kids take longer to get to the final point than other kids do. That's the observation. The question is -- and this perhaps fits with what David was just suggesting before, about the role of the family. The question is whether -- Checker -- whether you think that standards would be implemented in a more honest and genuine fashion, if in fact the consequences were put in the hands of the consumers, that is families, instead of the consequences in the hand of the government. So that if, in fact, whether it be vouchers, or parental choice, or something like that, you allowed the parents to pick up their kids and move them on the basis of whether schools were really at certain standards, whether in fact that would lead to a more honest and genuine notation of this.

MR. FINN: First, to comment on your comment. Every time I, myself, drift down the notion of breaking out of the prison of time and allowing time to vary, or holding the standards set, I am called up short. First, by E.B. Hirsch, with his insistence on grade-specific standards, and secondly, by an awful lot of voters and parents who want fourth grade to represent fourth grade, dammit, and they don't want any of this flexible talk about multi-age groupings, and stuff like that. So I think there's some implementation issues with letting time vary that are pretty significant. But I certainly agree with it. As far as the -- where the consequences come from, I've personally come to believe that we have the greatest promise of change if we allow them to come from two directions at once, which is why I believe that standards-based reform and market place-based reform are ultimately compatible, and indeed, complimentary, and maybe even co-dependent. That we are likeliest to get change if we have accountability upwards to the state standard-setters and testers and enforcers, and simultaneously, accountability downward to the consumers who have options and don't have to stay in a failing school if they see one they'd rather go to.

MS. RAVITCH: Bill?

MR. HONIG: Yes, two points. I think you have to distinguish different subjects. The time argument really does apply to something like algebra, because you -- algebra is legitimate.

There's a content in algebra that you should say every kid needs, or we want to maximize the number of kids that take that and pass that. But there are going to be differences in when they take it. That's completely different than third or fourth grade, when you want kids to read grade level material, and so forth, when you can say, "We want all of them to do this." So I think you have to distinguish between if you're going to go take AP math, you're going to have a different sequence of courses than if you just want this community college level. So that's one issue. The other one has to do with this argument that I wanted to clarify. When I was saying that algebra -- if you want to -- you want every kid to pass certain subjects, but the eighth grade is not a cut. It's basically, I want to maximize that, because that means I'm maximizing the number of kids who are going to take AP calculus, or get into these higher levels. It's a way of saying there's two levels. There's the basic level, but we also want to increase the amount of kids that make this higher level. And this argument that that doesn't work -- I think it does, because at least in the eighties we had this, "Let's max-out the number of kids going to college, or getting prepared for college, by fulfilling the requirements." Whether it was individually family-driven or school-driven, or some combination, those numbers went way up. And that was -- so the idea here is not to just have one measure, but a series of measures by which they can qualify in getting kids, more kids, to do this, or more kids to do that, or more kids to do that, won't just help on the margins, it helps quite dramatically. When you go from 30 to 60 percent passing a particular level, or 70 percent, that's major change in that score. That's not really marginal change.

MS. RAVITCH: Okay, I'm biasing my choices towards people who have not been able to speak yet. Richard Murnane?

MR. MURNANE: I just had a question for the panelists. If I understood correctly, one point on which they all agreed was that it did not make sense to have stakes for kids depend upon scores on tests taken at a single point in time. I wondered what they see as what ought to be done, and if they agree on that.

MS. RAVITCH: Checker, you want to -- quick answers.

MR. FINN: Well, certainly not a single test score taken once, and no other factors considered. That was my point. I think that test scores, plural, together with things like grades and teacher judgements and other things can -- ought to be weighed in terms of making high-stakes decisions about individual kids. It's like what a good college admissions office typically does, in terms of weighing a variety of factors about an individual person in terms of deciding they should get into that particular college. Now, this is real hard to do on a mass scale for statewide systems that are trying to be fair, uniform, and free of litigation. And to make these individual tailored decisions about individual people, very difficult to do.

MS. RAVITCH: Richard

MR. ROTHSTEIN: I want to emphasize again that we can have high-stakes, standardized, statewide tests to hold schools accountable. When I said before that we don't need to have those high stakes tests at the student level, somebody responded, "Well, kids need the incentives." But we've always had high-stakes tests at the classroom level. I mean, Albert Shanker always used to say that the American students' favorite sentence was, "Will it be on the test?" All of us

experienced high-stakes tests in school. Teachers are perfectly capable of constructing high-stakes tests for their students. If those exist simultaneously with high-stakes standardized tests at the school level, there's no need, and no reason, and no justification for trying to administer these standardized, statewide tests at the student level, because they give us very little valid information that we can't get by combining those other two types of information, the teacher evaluations based on their own classroom high-stakes tests, and the statewide evaluations, which tell us whether schools are functioning properly.

MS. RAVITCH: Bill?

MR. HONIG: Yes, I'll just add that I think I -- what does Checker say, mildly disagree? I think - I do not think that standardized tests for individuals are that important. But certificates are, end-of-grade courses are, and some outside external standards for those courses, like AP does, or international baccalaureate. That's the problem, is if you allow it just to be done at that particular school, you don't know what you have. And I think the resolution of your dilemma is just have a certificate where you can go after it once or twice or three times. So the standard stays the same, but you have multiple times to go get it. And I think that resolves some of the issues about high stakes with just one testing.

MS. RAVITCH: I'm going to ask John Bishop to speak very briefly, because this is the subject of his paper. Therefore, you can't give your whole presentation.

MR. J. BISHOP: The -- I've done a lot of thinking about this issue, how you impose stakes on kids. And I agree it's a tougher issue. But I also have a problem with the hegemony of the use of the word high-stakes. And we heard earlier, that standards-based reform is something that's hard to be against. But high-stakes tests is something hard to be for, and we've adopted this terminology. I mean, life is a high-stakes test. And the SAT is a high-stakes test. But a lot of what we're talking about, i.e. to pose high stakes on a school that is failing because their students are doing poorly, and that results in schools under registration, review in the state, meaning that some people come in and start looking over your shoulder, regarding what you're doing, and making suggestions about how you might change, and many calling into question the principal's job. And maybe we would want to find another principal, and maybe we might even decide that we need to hire some different teachers. I mean, that's life. I mean, that's applying to schools the kind of accountability and situation that most people live in all their life, in terms of their business and environment. And so to use a word like high-stakes to those types of consequences, I think somewhat exaggerates the degree to which -- unless we want to essentially say that we live in a high-stakes environment in almost all of our -- and that we simply generalize this word. Now, how to reduce the high stakes, how to handle the problem of putting stakes on kids. One is to have multiple opportunities to take the test. But I think a way of involving teacher input is -- has to be found. And in fact, it's done in other countries. For example, in Canada, where they have an end-of-course exam at the end of high school in most of the subjects in many provinces, it's only part of the final grade. And that's the primary use of the information, is that it goes into the final grade. That's the way the Regents worked, until it became compulsory, and you had to pass it in order to graduate. For 120 years or so, it operated as part of your final grade. And the fact that you got this credential rather than that credential meant you had some leg up. So it's not really high-stakes to get a Regents diploma, a local diploma, it's not really high-stakes if it's 10

percent, as is the case typically in Regents, of your final grade. It -- the key is that there are some states on an external assessment. It isn't -- we don't -- it isn't a linear relationship. The benefits don't accrue, become bigger and bigger, the higher the stakes are. You just need some stakes, not necessarily high-stakes.

MS. RAVITCH: Bruce MacLaury?

MR. MACLAURY: Thanks. A number of things that I was prepared to say have been said. But for Richard Rothstein, I wanted to propose a mantra, and that is, Compared to what? Compared to what, in the sense that your comments about Checker's paper, I think stimulated a great deal of good discussion. Should, can all students be expected to meet the standard? That generated a lot of good discussion. But in a sense, it seems to me the question is, compared to what we have, implicit standards as we have had right straight along. That is to say, kids do move from grade to grade by some criteria. They do, or do not get a high school degree, by some criteria. We've got standards. The question is that those standards are all over the map. And they may or may not mean anything. And what this discussion is about, it seems to me, is making implicit standards explicit. And yes they are, they are minimum standards, and that doesn't mean you don't have a higher option of going above them. You measure that by how far the scale of the ruler goes above the minimum standard. That's not a big deal. There is a minimum standard, and multiple attempts to beat that standard, it seems to me, is what we should be looking for. It's not sometimes practical to give multiple options to get to that minimum. And last comment is that yes, high-stakes tests, as John Bishop was saying, has a connotation that is tough, in effect. But what is the alternative? Compared to what? And compared to what is letting people slither through without any stakes at all. That seems, to me, to characterize more our current system than -- and so therefore, the question of high stakes has to be compared with no students, or no way.

MS. RAVITCH: I don't -- I have to, at this point, resort to let's get quick comments, because we're down to the last few minutes before lunch. And I know that everyone would like to get up and stretch their legs. So some people have not had an opportunity. Let me start with Uri Traisman.

MR. TRAISMAN: Yes, on this graduation test. If you don't know, the system that the courts upheld in Texas has two kinds of multiple options. It both allows students to take, repeatedly, the end-of-course test in three areas, or they can pass end-of-course examinations. So the state recognizes that some children are going to have special difficulty with cumulative examinations, and it allows them to take an exam in English and algebra, plus their choice of history or biology as one option -- and they can repeatedly take those exams -- or they can repeatedly take the three reading, math, writing exams. And that system has allowed lots of people to succeed, many of them, second, third, and fourth try. Different than teacher grading, which would have massive legal problems.

MS. RAVITCH: Lynn Olson, is that your placard up?

MS. OLSON: Yes, and this is a question that I don't know if anyone's going to be able to answer it this quickly. It seems to me this question of whether the assessments are aligned with the

standards is a key one that we do, in all discussions, tend to gloss over as sort of this black box. And the question I have is, does anybody have any ideas of how we get at that, other than leaving it up to the testing companies to tell us, yes, they are aligned.

MR. FINN: This is the reason achieve was created.

MS. RAVITCH: One of the reasons. Okay, hopefully you'll get an answer in private discussion, or in the rest of the day and tomorrow. Herb?

MR. WALBERG: This is a question, really, for the conference -- or these particular panelists, although it always --

MS. RAVITCH: This is Herbert Walberg, just for the reporter's --

MR. WALBERG: This is a general question, and I think National Assessment Governing Board was the pioneer in setting the standards, and now we all know that many states have an active standards board. Also, as Marci and Checker point out in their paper, many of the state standards really don't rise to the levels of the national assessment standards. And at the same time, some cities are considering having their own standards. For example, where I come from, Chicago, has been working very hard on working out standards that could be coordinated very carefully with Chicago goals and Chicago curriculum, materials, and so on. The question that arises, in my mind, is which groups would optimally set these standards, and how should they be related? That is, at the national level, the state level, and at the local level, and I think that's one question that is not, so far, coming out in the discussion or the papers very much. I'd like to see some of our discussion bear upon that question.

MS. RAVITCH: Alex Molnar?

MR. MOLNAR: Well, one issue that I'd like to see discussed that I haven't seen discussed is what bearing do these standards have on the structure of teachers' work in schools? Because it seems to me that on the one hand, we're discussing in this context, the creation of standards, the assessment of standards, and sanctions for not having achieved standards. But there's no conception of what does this mean for the teachers' work in schools, and how might that have to be changed, given any serious attempt for any conception of standards that one might deploy? So I don't think it will do to just lay this on top of the current structure of work in schools. That needs to be part of any discussion of standards.

MS. RAVITCH: I think that Bill Honig, you were -- this is what Bill is working on now, correct?

MR. HONIG: Yes, and I think your point is well taken, that you have to decide -- well, some of it's just obvious. If you've got a standard of taking more algebra, or getting every kids to pass algebra, you're going to have to change math all the way down the line if you're going to get those kids, and you have to think that way. If you're going to get kids to read fourth grade material by fourth grade, which a lot don't, you're going to have to change kindergarten, first grade, second grade, and people have to know enough. One of the deficiencies here is with just plain "there are consequences," is I know in San Francisco, and some other places where they said, "You're on the

list, and we're reconstituting the school." That's probably as strong as you can get for firing the teachers and moving them on. Even though that was said, and that was a potential consequence, nobody knew what to do. I mean, even though the pressure was there, they didn't know enough - this is just reading and math, now -- they didn't know what to do in that school, and nobody helped them do the specifics. So this is, I think, what you're getting at. You still need to have a pretty good agreement on what it's going to take you to get from here to here, and I think that agreement is around in reading and math. Maybe not in some of the other --

MR. MOLNAR: I'm saying the way work is constituted in schools.

MR. HONIG: Well, that's another issue.

MR. MOLNAR: That's the issue I'm raising. I'm making now, a flat-out assertion. There will be no standards with work in schools constituted the way it's currently constituted.

MR. HONIG: We have different experiences. We've worked with 20,000 teachers and about 1,000 schools, and most of them have improved, and most of them have, in fact, started to achieve this goal, and a lot of them did not change the work in school. The just worked in their classrooms and did better things. Now, I'm not discounting what you're saying, but if the idea that you just have to do that, or without that it won't work, I don't agree with.

MS. RAVITCH: Julian Betts?

MR. BETTS: I think Richard Rothstein raised an extremely important point criticizing simple pass/fail standards. At the same time, I think it's important we all understand that in a lot of states, we've gone beyond a single standard already. Twenty states have introduced advanced high school diplomas, and that's a way of differentiating and providing incentives to students at different ranges. And most states that have high school exit exams are pitching these at a grade 10 level of competence, or lower New York seems to be going backwards compared to other states that are adding additional types of diplomas. It's abolishing local diplomas. I'd actually like to know -- Rich has already made himself very clear on this -- I'd like to ask the other panelists how they feel about the abolishment of -- the abolition of local diplomas in New York.

MR. FINN: And their replacement biased the statewide mandatory regions.

MR. BETTS: Right.

MR. FINN: The -- incidentally, on your general point, the -- high school exit is only one of the places where this issue of multiple standards versus a single standard comes to bear. It also comes to bear for third graders, in terms of summer school and repeating grades, and stuff like that. So there's a whole slew of levels, and I don't know anybody who's got multiple third grade standards, and I'm not sure I'd be for them, actually. I think I wouldn't be. I think I'd be more for letting the time vary here. As far as the move to essentially standardize high school diplomas statewide, I think that is an inevitable consequence of statewide academic standards. There will be statewide reward or punishment, or credential or withholding of credential. It comes with statewide standards. I think that localities and individual schools ought to be free to supplement

the statewide standards in any ways they like. But if, in effect, you've said there are going to be statewide minimum academic standards in the state, then it's logical that they'd have to be met before a diploma could be earned.

MS. RAVITCH: Al Shanker deserves to be quoted here, because Al once said that any effort to create a universal standard is bound to produce a low standard. And I think that for high school graduation, that's probably right, and the story remains to be told as to what's going to happen in New York, in imposing a single standard for graduation for everyone. Sally, is that your --

MS. KILGORE: Yes, and I'll actually -- actually, I have a question and I'll actually say that Checker, you can answer this at any time in the next few days. I'd like for you to explain and justify the level of specificity that your particular -- I'll just give you three reasons, to give you some direction about why I think this is an important question. The certain level of specificity is going to lead to a certain procedural rationality, i.e. those outcomes that are steps toward reading, for instance, that maybe it's not embedded in the standards, you can have people meeting phonemic awareness, but still not understanding what their reading. There's another level where I think you can completely -- the expectations of understanding -- the level of specificity can be so great that grade by grade, you can't ever have the opportunity to allow time to vary, which I think is a critical thing if you want to have an equal standard. So sometime I would like you to justify what you think.

MR. FINN: Anything that starts off with "explain and justify" deserves time to get ready.

(Laughter.)

MR. FINN: And I've been given 24 hours to respond.

MS. RAVITCH: Okay, there was one other. Oh, yes, Bob Costrell.

MR. COSTRELL: I'm brief. In the same spirit as John's terminological point, I'd like to also make a terminological point, because I think it's important. And this is with regard to the argument against single assessments, or single test score, because that's a term that carries so much, it's even flipped Checker off balance here. What is meant by a single assessment? And I think it's a word game that's being played here, because it's not single versus multiple assessment that's at issue, it's external -- it's an external set of assessments versus local and possibly idiosyncratic set of assessments, which has all kind of problems that John's written about with regard to the incentives for -- and so on, and so forth. If you take -- and I don't think my state is alone, Massachusetts, I think a number of states have external assessments -- which they, in fact, will call a system of assessments, assessments system -- take Massachusetts. The first day, all the students in grades 4, 8, and 10 take a two-hour essay examine. In fact, it's not even two hour. It's untimed, okay? The teacher can come back and say, "If you need more time, you can take more time." It's untimed. And they are told to draft it and revise it, okay, a kind of activity that they don't do in their 50-minute class. All right? Then there's another part of the exam, another day, you have another series of days of testing, there's an open response section. There's a multiple choice section. So the idea of having multiple assessments is that you can do better on one part, say the essay part, that will offset doing worse on the other part, such as multiple choice

part, okay? So you know, where is the boundary around the term assessment, when you say single assessment? I don't think that's really been thought through with the kind of knee-jerk use of the term, single assessment, which is a very powerful rhetorical device against these exams, when they are, in fact, multiple assessments. The real issue is an external set of assessments, versus local and possibly, as I would say, idiosyncratic set of -- and this also gets back to the majority of the use of the term standardized.

Searching for Indirect Evidence for the Effects of State-Wide Reform
Authors: David Grissmer and Ann Flanagan
Comments: Uri Traisman *and* Rob Meyer

GENERAL DISCUSSION

MR. RECKASE: Yes, one of the things that I have heard recently, and maybe someone else can verify this for me -- was that in -- it's either Connecticut or Massachusetts -- that they went out of their way to make their state assessment, in writing, look very much like NAEP writing, and that they've had big gains in performance on NAEP, because essentially the kids are practicing on a test that looks very much like NAEP. So it made me wonder when I was listening to this about the differences across states, about how similar the tests in Texas are to the NAEP tests, and how similar the tests in California are to the NAEP tests. You know, are there effects that are in here that are related to the similarity of the state assessment, or the high stakes assessment that the students are taking within their state to the test that you're using as a criterion for looking at improvement, and if there's any, you know, information on that.

MR. COSTRELL: Well, I can answer directly for Massachusetts, which is that the -- Massachusetts, the MCAS, that's what you're referring to, only began to be administered in 1998, so it could not --

MR. RECKASE: Okay, maybe it was Connecticut then. I knew it was when the -

MR. GRISSMER: No, Connecticut -- North Carolina did test with NAEP -- Texas, I think Texas didn't even try to do anything with NAEP.

MR. MEYER: It's a matter of principle.

(Laughter.)

MR. GRISSMER: And it's clear, I think, you can get a one-time gain from not alignment to alignment. North Carolina did it in the late eighties, when they designed their tests. They basically modeled it on the NAEP test. So the 1990 to 1996 results, I don't think -- they might have reflected a little of that one-time gain, but I don't think a lot of it. Connecticut is a more interesting case, and it's possible that at one time -- the blew the socks off the fourth grade reading test in 1998 -- I've seen a report on that, and aligning with the NAEP may be part of it. But I think there are other reasons in Connecticut, as well. But it's hard to say.

MR. TRAISMAN: In Texas, well, it historically it's not the case that people looked at their NAEP criteria. But a very rich source of data is that almost all districts in Texas use a commercial test in addition to the state test. Forty of the fifty largest districts use another test. So the state samples -- you can see two things from this. Newly higher results in the state on the commercial tests that are very different characters, SAT-9, Metropolitan. But really interesting to us is there are some schools -- and you can pick them out -- where anything not on the state tests, the kids do zero at. Basically, you could see the effects of teaching to what I think is a broader test. Then we have a whole collection of schools where the learning is very broad. So the holy

grail here is, what is it about community and school leadership that allow schools to use the state test as a platform, rather than a ceiling? That's the kind of practical question that a guy like me wants to know.

MS. RAVITCH: Gerunda Hughes? I'm not close enough to read your placard, but I think -- pardon me if I mispronounce your name.

MS. HUGHES: Uri, I was very happy to hear that the amount of variance that's explained by family characteristics was reduced dramatically from about 60, 70 percent to 30 percent. And the question that I have is that I'd like you to respond to the allegation that perhaps it had to do with the restriction of the range of family characteristics, and that over time, many of the reform efforts in the State of Texas has led to the dropout rates increasing, pushout rates increasing among African-American and Hispanic students. So in fact, what we see is reduction in variability in that group, in terms of family characteristics. And if that is so, what can we do to maintain the variability, and at the same time reduce the amount of variances explained by family characteristics?

MR. TRAISMAN: That's -- you know, of course these are central questions. In the fourth, fifth grade data, we have -- I don't think anybody believes that there's significant dropout at that level. Our look -- and this came up in the court case -- there was very little evidence, in fact, no evidence that the state accountability system was leading to dropout, even at the eighth grade level. But there is a piece of data that is stunning. If you look at the NCES data on dropout, the proxy is being overage for your cohort, all right, which is highly correlated with dropout. In most states, the data is a little bell curve there, rising and low in the seventh, eighth grade, ninth and tenth it's high, and then most twelfth graders graduate. In Texas, it's all concentrated in the ninth grade. There's almost no dropout. Once kids get to the tenth grade, and they get to the point where they have to take the exit test, they graduate. The one year where there's no testing, the schools are pushing kids out in the one place where there's no light shining on them. Schools are magnificent devices for pushing kids out they don't want or can't deal with. The legislature, because of the richness of the data, was able to see this, and explicitly address it in legislation, changing how the accountability system works and also focusing significant resources on that direct dropout. So I don't think there's any evidence -- we haven't found any, research studies haven't found any -- of K-8, these things being due to fourth graders, fifth graders, eighth graders dropping out. There is a problem at the ninth grade, but even then, Texas graduation rates are pretty high, nationally. Dropout rates are pretty low. So we're not seeing it, but it remains a worry, and it's still the case that far too many kids are dropping out.

MR. FULLER: And since 1994, when the current accountability system was put into place, the actual number of students who were progressing from ninth grade to graduation in four years has increased.

MR. TRAISMAN: Increased.

MR. FULLER: And that's contrary to what some people are putting out there. And Martin Carnoy from Stanford independently -- he was doing a study in Texas, and he found the same thing that we did, that that's actually the -- staying in school and graduating, that rate, is actually

increasing, has increased, since 1994.

MS. RAVITCH: Alex Molnar?

MR. MOLNAR: I want to come back to the differential in math achievement and reading achievement.

MS. RAVITCH: Speak up, because they can't hear you.

MR. MOLNAR: I'm so seldom asked that. Thank you.

(Laughter.)

MR. MOLNAR: I want to come back to the differential and achievement between reading and math, because if you take a look at the various analyses of the voucher achievement data in Milwaukee, two out of the three of those data show that there is a gain in math achievement, at least arguably. Two out of the three show no particular gain in reading. If I take a look at our data and our evaluation of Wisconsin's class size reduction program, the SAGE program, we see larger gains in math than in reading, although the differential still remains significant. And if I understood you correctly with regard to the math instructional methods in Texas, they are varied. I mean, you have quite a diverse array of instructional methods, so you can arrive at the standard in a multiplicity of ways. I would assume the same is true with regard to reading. So it doesn't -- so in reading you can only arrive at the standard in one way?

MR. TRAISMAN: No, no. It's not a question that the standards -- the way in which the professional support, capacity-building was organized --

MR. MOLNAR: Okay, that -- I would like to hear more about that because I want to try and understand more clearly your hypothesis in that regard, and I want to try and think it through in relation to what our data in Wisconsin are showing, with regard to the impact of class size reduction and what Bill said earlier about the work that he is doing in California with regard to capacity-building among teachers in the instruction of reading.

MR. TRAISMAN: It's a long discussion. I'm not sure --

MR. MOLNAR: Well, I said more, I didn't say everything, Uri.

MR. TRAISMAN: It may, first of all, be that it's harder to make A's in reading. We can't counter that, or address that. But the point is that the -- so the policy frameworks for reading and math are the same. But the way in which the state supports schools meeting the standards is completely different, all right? The group that runs math has taken a different approach, aimed at local control, pluralism, continuous improvement ideas. So that instead of having a uniform state program which puts out a preferred pedagogy, the math approach is to focus on looking at student data, looking at your own student work, looking at the textbooks that sometimes you, as a school, have little choice in adopting, often it was a political district decision, and working from what you have. It's a bottom-up approach. Analyze students who move that way. You don't

have battles between the Saxon folks and the NCTM folks. They actually reasonably well collaborate. The algebra standards got a universal approval by a very contentious state board. So there's not the same sense of politics around it. There's more of a focus on student achievement. There's a different feeling around reading, all right? There's a uniform state approach to train reading teachers. It's not a bad one. I, you know, would think it's pretty solid for a research base, but it's not perceived that way locally. People have very strong feelings about it. And I'm speaking perception. A perception is that lots of people are withdrawing their energy from reading, being involved in the improvement of reading. They have to choose sides. You don't have to choose sides in math. The cultures of reading instruction and math instruction are very different in the state. And I'm saying that some of these gains may be attributable to that, and it's worth studying the nature of support for, you know, for the implementation of the standards.

MR. MOLNAR: Yes, and bringing back -- if I understand you, that has implications for our discussion before lunch, with regards to where standards come from, who sets them, and so on, and the implications in the classroom.

MR. TRAISMAN: Right.

MR. MOLNAR: Okay.

MS. RAVITCH: Okay, Ed Haertel.

MR. HAERTEL: Thanks. I also wanted to speak to Uri's observations about the difference in the -- the striking differences in the changes at the school level between reading and mathematics over time. And your comments just now, repeating the differences and the cultures, a lot of improvement in those areas may reinforce the hypothesis that's implicit in the question I want to ask. One take on that difference is that mathematics is a subject in which it is easier to teach than to test, because it is a subject in which there are a large number of discreet things which one needs to learn. And if one has learned to do the sorts of problems that appear on the test, he has learned those to recognize how to apply rules in the particular kinds of situations that show up on the test, then one will do better. With reading there are a few things we can do to teach to the test. We can focus on answering brief, actual questions, we can read shorter passages instead of longer passages, maybe we can tweak the genre, we can stop wasting time on things like class discussion and analysis. But in mathematics, I think that especially if one starts from the bottom up, as you say, looking at this data, it's possible to get a bigger bang for the buck on tweaking instruction to get the test scores up.

MR. TRAISMAN: I think that's true to some extent. If you look at the tests, though, which we could take on line -- and you can see them, they're released immediately after -- you can see that there are a substantial number of what I think you'd call conceptual questions. They're not questions where you have to just pick an algorithm. I think the mathematics is much harder than people think it is. People think it's just a set of 120 algorithms or problem types you have to learn. That's not the way the Texas standards are organized. Just a comment about it. In the focus group mechanisms for developing the standards in math, there were people who felt very strongly about memorization of facts, and kids knowing particular things. And there were other folks who felt very strongly about some abstract idea of understanding and being able to put the

math to use. What we found in the groups, borrowing from public agenda and some others, using their methods, is that the people who talked about understanding had ideas about knowing facts. But they also talked in terms of this math as cultural property. They -- we have tapes where they talk about remembering doing it with their mothers or their fathers. Very important. So when we built the standards and the tests, which by law are based on the standards, we built standards around knowing facts and recall, and we also built explicitly standards around demonstrating and understanding of ideas. So we didn't see these as competing. So we might be able to do -- it would be interesting -- to do an analysis of the state math tests and see on which kinds of items the gains are largest. The state data source is so rich that it's possible to really, I think, get somewhere with that question.

MS. RAVITCH: There's actually a very simple answer to your question, which is that clearly Uri wrote the math standards.

(Laughter.)

MS. RAVITCH: Who wrote the reading standards? There would have been gains on the reading standards, as well.

MR. TRAISMAN: You let that one get away from you.

MR. BETTS: Okay. I'd like to ask David about consistency of the NAEP scores over time. If you look at the official NAEP documents, especially in the early nineties, often they declined to actually report whether gains or declines were statistically significant, because of differences in the sample. What's going on there? Do you think that your controls have controlled for that adequately? And what are the lessons for the rest of us who use these things in a much more casual way than you do?

MR. GRISSMER: We were correcting the 1990 and 1992 state scores as late as 1997, due to an error in the program here and there at ETS. So you've got to keep your eye on the ball when you analyze state scores. I was, I think, flabbergasted by analysis of state scores. There are a set of 271 scores over time. I thought this was going to be a sort of fun kind of little exercise in -- but nothing serious. And in the process, I have learned more methodology, a respect for small sample sizes that had good T values, which I really didn't have. I mean, T values do make a difference. If you get a coin and flip it 10 times and it's heads, that's as good as, you know, 150 times. That is -- but you have to trust your T value when it comes to small sample sizes, as long as you do the sensitivity analysis, which for small sample sizes are more -- But we get results which can predict the Tennessee experiment. And the sense I have about -- I think there's still sort of -- you know, I think what you're referring to is Department of Education only compares, you know, this year's eighth grade test against two years ago, and they put a star on it. Now, what we tried to do is have more consistency. A state, to do well, has to get sort of stars on every test and show us a consistent gain upward, over a time period. And so it's more of a robust test, in a sense, than the individual NAEP test. But I've come away from the -- analyzing this NAEP data with a very healthy respect for the test, a sense that, you know, ETS does a really pretty good job of the -- all that good spiraling, and all that stuff they do, and that when we start to analyze the individual level data and play with the different quality of family characteristics, we're going to learn a lot

about the methodology. For instance, Rob -- I mean, a year and a half ago I would have started the analysis that Rob suggested, fourth minus eighth, 1992, 1996. I think Tennessee really puts a hole in that. I mean, I don't think we can do that kind of thing any more. The resources that go into a kid's eighth grade score have as much to do with what happened in first and second grade -- and it's not always captured by the fourth-grade score. So that it really is a serious methodological issue on these. But I didn't start off as a respector of this analysis, and I tried for 225 pages to explain why I did --

(Laughter.)

MS. RAVITCH: Bruce MacLaury?

MR. MACLAURY: Yes, briefly, I wanted to ask -- you intrigue me by stating that the -- politically, the state started out with relatively lower standards, and then rached them up. My question is, is it really the content standards changed, or that the bar against those changed over time?

MR. TRAISSMAN: Both. The content standards have now gone through an iteration. There were complicated political discussions about how much more rigorous to make it this time, versus five years out, when it's done again. And then the bar was set very, very low, 25 percent passage rates for each group. And legislators, under the umbrella of a powerful business group, basically sanctioned that, and took a hell of a lot of heat for doing it. They understood -- legislators understood that they couldn't make a system in which most of their constituents failed. And the opponents of public education were pushing very, very hard to make the standards high enough so that the system would collapse. And the business community saw that strategy for what it was, and addressed it.

MR. MACLAURY: Was there an announcement at the outset that the standards would rise over time?

MR. TRAISSMAN: Yes, but not that they were being set low, initially.

MR. FULLER: The actual curriculum standards however, they were adopted in 1985, the essential elements, and they remained in existence until 1997, 1998. I think in 1998, the Texas essential knowledge and skills replaced those. What did change is we started standardized testing in 1986, with the TABS, which is Texas Assessment of Basic Skills, and then we replaced that with TEAMS, which was a little bit harder than TABS, and then replaced it with TOSS, which was harder than TEAMS. So it's a really ingenious strategy. And then we're now aligning it entirely with the NAEP curriculum standards. Right now we're testing to overlap between the old curriculum standards and the new curriculum standards.

MR. TRAISSMAN: Just one comment that's really important. For new items, it takes teachers about three years before half their class can do the item. So what you see when you look very closely, is something that is the most important, is that it's not the kids. And you can actually see that it's not the kids.

MS. RAVITCH: Richard Rothstein?

MR. ROTHSTEIN: Just a quick question. The analysis you've put up is an analysis of scale scores, right?

MR. TRAISMAN: Yes.

MR. ROTHSTEIN: It was not percent passing? So how -- if you did an analysis of the extent to which the variation attributable to family was reduced in terms of percent passing versus scale scores, how would the -- what would be the significance of that?

MR. TRAINMAN: Yes, we've done -- let me turn to Ed Fuller, my trusty -- we played around with this.

MR. FULLER: I believe there is -- it essentially looks the same, although I haven't run an analysis, and somebody else has run that analysis. And it essentially -- the correlations look fairly similar, with a little bit of difference. And I know that in -- the mathematics definitely looks the same. I haven't seen the reading. I can do that, and we'll share that with you.

MS. RAVITCH: Larry Steinberg?

MR. STEINBERG: I have a question on math and reading issues. That work paper says you're doing a summary, not just -- you do a whole bunch of tests. But if I count right, it's -- many of the comparisons are with the math, and only one used the name reading. And I think your cutoff is 1990, 1996. You have anything since 1996? I wondered if you had a chance to look at the 1998 -- and add that to the analysis, and does that change any of the results? Does it still hold up?

MR. GRISSMER: The analysis I presented today was five math scores, three eighth grade and two fourth grade. We'll have to weed scores out, because it was just in 1992 and 1994. And I think, talking about the quality of the NAEP, the one area that I think -- the 1994 test is actually -- kids scored a little lower than 1992, two years later, which I think probably is a -- problem, more than anything else. Yes, we have analyzed the 1998 reading scores with the 1992, 1994, 1998. Preliminary results show that, you know, the gains in reading, compared to math were about a third as large. In terms of state sort of rankings, Texas falls a little, for the reasons that I think he's -- I mean, they fall from second to fifth in the ranking. Connecticut and Kentucky and a couple of other states who really did well on that test come out -- But the movement is not -- I mean, the correlation, I think, was 8.5 in the rankings in the reading and math -- they're different -- while the gains were different, both the value-added and the ranking of the gains were -- my impression is were not markedly different. You could pick out states that did differently, but we're not talking about two really different kinds of results.

MR. STEINBERG: What about the effect of all that on your conclusions about how the standards reform is affected?

MR. GRISSMER: I think it does rank more on math. And I -- what I was talking about in

Texas, I think, is a nationwide kind of thing. I mean, I -- and we haven't been as successful with reading as we have in math, nationwide, I mean, I -- for that states in the sample. And particularly in the central cities, the reading scores where you'd expect the most trouble really have almost declined a little. And I think for the reform agenda, I don't understand the reason, but I mean, we're -- and Texas has one. There certainly is evidence to me that it's moving math much more than reading.

MS. RAVITCH: There's a question here, but I can't read your placard because of your --

MS. PHILLIPS: I'm Meredith Phillips. My questions sort of are --

MS. RAVITCH: I'm sorry, what's your name?

MS. PHILLIPS: Meredith Phillips.

MS. RAVITCH: Oh, hi Meredith, how are you?

MS. PHILLIPS: Hi.

MS. RAVITCH: I still can't see her.

(Laughter.)

MS. PHILLIPS: I want to posit sort of an alternative explanation, maybe, for this reading/math difference that I can see raised kind of a -- to you, and that's basically the math that's taught in school did an awful lot of overlap than the math that's taught at home, at least during the school year. So you learn some math at school, and maybe take some homework home, but you don't prepare -- but they do teach you a lot of verbal skills, sort of indirectly, by speaking to you, and they may -- you may yourself, choose to read at home, so that you would expect that the reading scores would continue to be more highly related to the family background than math scores, you know, across the board. And also, that there's this question of summer loss. There's a lot more summer loss, we think we know, during -- in verbal skills than in math skills, that essentially you -- and it's differentials, that the black/white gap, and the gap between poor kids and rich kids increases more in reading and sort of vocabulary skills during summer than during -- than in math, and then during the winter. So you might imagine that you basically have sort of a widening gap, because of what's going on in the summer, unless you actually tested kids in the fall and the spring, and sort of measured gains that way. But you're always going to see sort of a strong correlation between verbal stuff and family background. You know, always, but more especially unless you -- does that seem reasonable?

MR. TRAISMAN: I wonder how that --

MR. GRISSMER: I think -- I mean, the summer effect is a really good -- it's pretty well documented, and I don't know the mechanism, necessarily, of why you forget more reading than math during the summer, but that's what the results say.

MS. PHILLIPS: Well, it's not just that you forget -- well, it's actually -- if you look at sort of most of the data, it looks like rich kids gain more verbal skills and poorer kids flatten out. And so essentially they're supposed to be more verbally -- a verbal environment during the summer. But there's also this issue that we assume that what goes on during the school year is just a school effect. But in fact, you go to school at 8:00, come home at 3:00, and there's something goes on in your neighborhood and in your environment, that presumably is more -- would affect your verbal skills more than your math skills.

MR. GRISSMER: Yes, I think that's a really viable --

MS. RAVITCH: Ron Ferguson?

MR. GRISSMER: I think Ron Ferguson is going to --

MS. RAVITCH: That's going to have to be the last comment, because we're just about exhausted our time here.

MR. FERGUSON: I just want to posit an alternative interpretation for your point with regard to the Tennessee STAR experiment, when you say the kids who are only in for a year or two lose their gains by middle school, whereas the kids who were in for four years don't. One thing that's different is the kids who were in for a year or two probably move around a lot more, or at least somewhat more than the kids who don't. We know that mobility tends to interfere with sustained gains also. So could that differential have something to do with mobility and so you've got to omit a variable problem, rather than the notion that the earlier test score doesn't capture much of the -- doesn't capture all the base line.

MR. GRISSMER: That's possible, sure. The ultimate explanation that I have is more of a resiliency one, but I think there's other literature that says the duration of the effect is sort of proportional to the duration of the intervention. A one-year intervention is not likely to last a long time. A two-year will last longer.

MR. FERGUSON: And I want to believe that, too.

(Laughter.)

MR. GRISSMER: I think -- you know, I forget -- I reviewed the paper, but I now forget whether that was controlled for the specific analysis that Larry did or not. There was two papers there, and that's a really good point. I don't know.

MS. RAVITCH: Okay. I think this is a good point at which to stop and change players. And thank you very much to the panel.

The Controversy Over the National Standards
Set by the National Assessment Governing Board
Author: Mark Reckase
Comments: Michael Feuer *and* Ed Haertel

GENERAL DISCUSSION

MR. MEYER: I wanted to get at the conceptual motivation for this standards-setting, and let me talk about an alternative. Suppose that you had a bunch of teachers, experts, say you know, "I know what's going on in the world. I look at people, and I think that it'd be good for people to be advanced, and I kind of define that cutpoint of advanced is it would be the top five percent of the people in the country, and similarly." Now what you have there is essentially setting cutpoints through norm reference. And I'm wondering -- here's the -- why isn't that a good way to go, compared to this other process?

MR. RECKASE: Well, that one's a tough one to answer, because it's really a NAGB policy issue. They wanted a criteria and reference standard. And that's the way the original policy statement was produced back in 1990 was to say, "We're going to set criteria and reference standards -- this is a standard for what students should be able to do, not necessarily what they can do at this particular point in time."

MR. MEYER: Well, but --

MR. RECKASE: Now, I'm maybe not the person to answer it beyond that, because --

MR. MEYER: Let me just -- why couldn't they have said that, you know, their standard is they want most -- they want people to be like some sense -- some people they know who are -- and they would identify those people as being at the 95th percentile. I mean, it's not clear that that's different.

MS. RAVITCH: Checker? And you were there --

MR. FINN: Yes, there's a bit of history that's worth lying on the table. When the national goals were set in 1989, goal three used the phrase, "Students shall demonstrate competency over challenging subject matter." Students shall -- all students -- shall demonstrate competency over challenging subject matter. And then the governors proceeded to name five subjects in which this competency over challenging subject matter was supposed to be demonstrated. The governors, in the national goals, did not however, define what competency over challenging subject matter meant. NAGB, unbidden, shouldered the responsibility of trying to spell out what competency over challenging subject matter would mean in those subjects, and that was the proficient level on the three levels. And in fact, embedded in something that Mark flashed very quickly on the screen as he was showing the sort of once and future definition of proficient, embedded in the definition of proficient still is the phrase competency over challenging subject matter. This was directly taken from the national goals as set by the governors, and was meant to be a judgement call of a criterion referenced sort that would equal the proficient standard, below which there would be a basic standard and above which there would be an advanced standard. But that's what it was

meant to do.

MS. KILGORE: Would you give me two minutes?

MS. RAVITCH: Sally, go ahead, but --

MS. KILGORE: I think that the whole standards movement is premised on a move away from percentile -- which has not been very apparent from the discussion today. And it's meant to suggest that students -- having a distribution suggests that we only care about rank order, having a mastery of proficiency is distribution -- all children could be proficient, all children could be inadequate, all children could be advanced, simply by virtue of the fact that they had mastered a certain amount of knowledge, not because they are at some rank relative to other people. So that the criterion reference test moves it away from distribution --

MS. RAVITCH: Okay, Ed Haertel wants to say something --

MS. KILGORE: Good.

MS. RAVITCH: -- to this one.

MR. HAERTEL: Picking a given cutpoint, or given -- or statistic distribution at a given point in time is not the same thing as saying we're only interested in percentile ranks. This is what was done for TIMSS, it's been suggested for NAEP, it could be done, it's an alternative way to benchmark.

MS. RAVITCH: Bill Honig?

MR. HONIG: Yes. I think I'm agreeing with Ed here, but I'm not sure, so I'm going to put this out. It's a comment. I think one of the problems in the NAGB is that they're using this complicated procedure with items and so forth in places you don't necessarily have to do that. I think you're being very sanguine about bringing experts together and have them get the right idea. You can bring committees together forever and they still blow it. So there's got to be some other check. I'm going to give you an example of that. In -- Diane's working on this now, and -- talked about it -- in setting reading standards, which is a little bit different from the other ones -- but setting reading standards, there is a content validity issue. One of the things you want is can a fifth grader read fifth-grade material -- or, in this case, a fourth grader read fourth-grade material -- and understand it generally? The bias of the group was the more esoteric thinking skills. So they left out this kind of base, which is really important. And so one of the issues in reading is how far up that scale you can rate them. And you can get, very easily, comparable passages that are at fourth grade and see if kids can understand them and set up a standard of 75 percent, or whatever the comprehension standard -- and get pretty close to that. That is one part that's missing, just as in math, they missed some of this basic understanding, and a lot of it's more esoteric. The open-ended questions in reading really test writing and some of these more complex skills. And that's part of the game, obviously. But it leaves out a major part, which is, "I want to get more kids to read this, be able to handle this stuff that the teacher is giving." Now, NAEP gave a great test -- when was that? In whatever it is, they did a fluency test where they gave a

sample, and they -- and you had to get -- they measured accuracy, and they measured could you make the right intonation, and they measured speed. And they found out they were highly correlated. But that testing procedure was a lot different than items. That was basically an individual given, "I've got to see how the kids does." And I think that would help with some of these validity issues, is to try and think about what you're trying to measure, what the real world is, why you're doing it, and then developing procedures that are more in line with that, instead of this one way fits all, which works some times, but not other.

MS. RAVITCH: Matt Gandal?

MR. GANDAL: Thanks, Diane. I'm interested in lessons for states here, talking about the governors setting the goals in 1989. There are a lot of governors struggling with how to do this themselves right now in their states. And one thing that I took away from -- it seems like everyone on the panel agrees that the motivational factor is key here, and that that lack of motivation in NAEP makes it very difficult to interpret results in certain ways. It brings me back to the earlier conversation, when folks around this table were talking about high-stakes for kids, or attaching consequences for students, and whether or not that was the right thing to do. I'm taking away from this conversation that you would suggest to states that they think long and hard before not attaching consequences for kids to their assessments, if they actually want to make sure that all the people in the schools, including the students, are trying as hard as they need to, and they could really interpret the results in the best possible way. So states like Kentucky, and others that have not done that, might want to think a little bit differently. That's one question. The second question has to do with looking globally. I've always been fascinated how countries like France and Germany and Japan, that have very high-stakes tests, that do have a combination of multiple choice and performance items that are very challenging and again, they matter for kids, seem to get through all this over the years without these big debates and fights about how they set cutpoints. And in fact, having looked at the tests in the earlier days when I actually worked with Al Shanker, it became very clear that these are challenging, but you don't hear all these struggles about where to place the barn. I'm wondering, in the research done for NAGB or the studies, Mike, that you did to look at the NAGB process, whether anyone looked at -- across the seas at how this is done elsewhere.

MR. J. BISHOP: You don't live in the countries. It goes on there, too. Living there -- after these exams are given, they are released, and there's a discussion on television by experts about whether the question was more or less difficult than last year, and how appropriate it is, and what's the right answer, and so forth. It's central to the French way of thinking about things. If this -- people fight about this kind of thing always.

MS. RAVITCH: Well, that's very encouraging to hear, but --

(Laughter.)

MS. RAVITCH: Roy Truby:

MR. TRUBY: Just one comment.

MS. RAVITCH: Roy's executive director of NAEP.

MR. TRUBY: Just one comment. You know, the policy has changed quite a bit over the 10 years, and I think that many of the board members felt that if you looked at this more holistically, and judges saw the entire book, that that might change things. We also reluctantly brought in consequence data into the achievement level-setting process, and brought it in much earlier than some of the previous boards, I think, would have agreed to. And many thought that that would have the effect of lowering the standards. And in fact, it didn't. When you look across subjects -- and Mark, you could correct me if I'm wrong, I'll be real quick -- generally whether civics, or math, or reading, if you look at the total points available on the test, the cutpoint for basic generally means that they get 40 to 50 percent of the test right to be at the basic level. And at the proficient level, it'll range anywhere from 70 to 82, or something, you know, they have to get about 70 percent of the test right to be at proficient. And generally, they have to get about 85 percent of the test right to be at advanced. And when they look at that whole booklet, and they'll say, "Well, this test wasn't that hard, and you know, we don't think 70 percent of the possible points on that test is unreasonable to be at the proficient level," and then they look at the consequence data and it doesn't change very much, their original take on it. Maybe you would comment on that, Mark.

MR. RECKASE: Right. The four rounds of ratings that they go through, what the typical result -- you do get slight changes up and down, but not a lot. It's not dramatic as they get more information. It does get more consistent, it pulls in the outliers, the people who were extreme tend to get pulled in and there's more agreement, more consensus, as it goes along. But the -- it is an amazing -- and then the consequences data have been moved earlier into the process, and the people, you know, very seriously take the charge that they're looking achievement level descriptions and they're translating those descriptions into cutscores, and that that's their job. And that's what they do. And then when they see the consequences data, they say, "Well, I wish it would be better, but to go over and change the result would make it inconsistent with the description." And so there is sort of a -- you know, they stick with their guns when they get to the end of the process, and they don't change things.

MS. RAVITCH: Excuse me. At the very end, I'm sorry, I can't read the placard. Sylvia Johnson?

MS. JOHNSON: Sylvia Johnson, yes, thanks. I was just going to say, to add, that we do have procedures for empirical studies with real student data -- they call them norms -- but they're not appropriate, as has been discussed for -- even in terms of wanting a content and process referenced instrument. But it seems as though achievement levels could be structured that would include both sorts of data. I think part of our problem is a subjunctive problem, in that, you know, do we mean should because of what has happened to students in terms of their actual experiences, or do we mean should in the sense of the Latin owe, ought, must, because they must be able to do this as a result of what -- of being in the fourth grade. And I think that the -- it tends to be more of the latter, rather than looking at what sorts of experiences the children have actually had, and even what sort of testing-taking experience they have. I don't think we've looked enough at the real life situation of testing, or the real life situation of learning in the classroom for the majority of our children. We have -- just that a validity studies panel has

released a report recently showing the great range, in terms of testing-taking conditions. It was done by AIR with Stan Kavage, a couple of other authors, showing that higher scoring schools had very different styles in terms of test administration, the number of children that were at desks or tables, the whole atmosphere, the noise, the disruptions, and you know, that can have a great deal to do if you have three or four items in a block that are missed, you have a great deal of change in terms of performance on the block. So I think that we do need to look at real student data, but look at it in terms of what is actually happening to children as we move along with trying to set achievement levels based on standards that are defined in terms of content and process.

MS. RAVITCH: Okay, Larry Steinberg?

MR. STEINBERG: I want to go back to the motivation issues, here, and just with one perhaps perverse observation. And that's that it doesn't strike me as necessarily the case that evaluating people when they're running for their life is a more valid way to assess what they know or can do than evaluating them when they're jogging. As someone who has taught college students at big public universities for 25 years, I would say that most of the students are jogging, rather than running for their lives. And that's certainly what we find when we look at kids in American high schools. So it seems to me that administering a test under the motivational circumstances that are probably more realistic than the motivational circumstances behind the SAT, would give us a more accurate gauge of what American students can do than what the SAT does.

MS. RAVITCH: Two more comments, and then we're going to take a short break and then come back for John Bishop's paper. Herb Walberg and then Maris Vinovskis.

MR. WALBERG: I'd have to agree with Michael Feuer, that the NAGB achievement levels have probably been criticized more than any other piece of education research or assessment than we've ever had in the United States, and it was particularly gratifying, as a founding board member, and chairman of the design and analysis committee, to hear David say it's the best process that's ever been implemented in this field, even when compared to law and medicine. I was also even glad to hear Ed's comments, which I think were somewhat critical, but it exemplifies what the board tried to do. It's own board members were sometimes much more critical even than the outside critics. So I attribute the continuing success to people like Diane and Roy and Checker who maintain this and maintain the process that Congress wanted to do, even though there was such a storm of criticism. I would like to also -- to return to the question of motivation. And I would like to say that the board was well aware of this even when we first started it, and we tried to get the National Center for Educational Statistics to do some serious studies of this. I think a couple small studies were done, but we had a presentation before the board, even in the second year, where we psychologists, like randomized field trials, where there was random assignment to sort of more optimal conditions, that "You're doing well for your school, and you've just got to count on your grades," as opposed to rather neutral instructions that were given to others, and we found an effect size of about two-tenths of the standard deviation, which is not large, but it's not trivial, it's a kind of a moderate effect. I'd like to call attention to one thing that I would love to see the National Center do, is more systematic analysis of this, because we've heard so much speculation about it. I think, too, that if this were to be done, NAEP could even be more useful, because I think one of the original purposes, or at least we saw the standards were that they could be

calibrated with states. And if we could make an adjustment for the motivation effect, we could more readily be able to compare say, Illinois with Indiana.

MS. RAVITCH: Maris?

MR. VINOVSIS: I really enjoyed Mark's paper, and I thought it really gave a good sense of the overview on some of the debates. But I wonder, Mark, in your revisions, whether you wouldn't contextualize, because there are other controversies that maybe equally are more important than the ones you've covered. And I would want to remind myself that isn't it interesting that the congress got so involved in these studies on such technical issues that have great importance, and we have GAO studies, everything -- so much work, as you say. On the other hand, we have things like prospects, marginatudinal study of school change, perform works, hardly any interest, hardly any work done on it. We have Porter passing comprehensive school systems, which is supposedly research based. Hardly any interest. Why the disparity? I would suggest to you that when you rewrite the paper, you might want to contextualize it a little more in terms of what's going on in the early nineties, when this is happening. This is a situation, for example, where people are fighting other battles. Some people are trying to get rid of NAGB. And the paranoia, and it's -- again, I'm not saying there's no basis for it -- was immense in terms of the House, where people have said, "Let's give this to NCS, let's not have NAGB do it," and trying to show how bad NAGB is, rather than thinking that it's an open process -- and I'm not saying it's not, it's certainly open in retrospect -- but the stuff will be in controversy and attempts to do the contracting certainly fed that paranoia on some people's parts. Many of the scholars and people who got involved in this also didn't like the idea of these testing other kinds of things. So my suggestion to you is you'll do us all a favor by trying to think about that context, because it's not just the issue of the standards themselves, and how hard it is to do it -- which I think you do very well -- but by giving us that context, I think people who are not as familiar with this thing will benefit, because part of that is very specific to the times. It may not be as much that way in the future.

MS. RAVITCH: Sixty seconds for -- last comment from Mark.

MR. FEUER: I just want to endorse the idea of context here, and one of the things I didn't get to say was that a good reason to worry about the credibility of these achievement levels is because of the reaction that they engender about the underlying origins of this entire exercise. Now, this is something I have to say very delicately and carefully. This is not my opinion, but there are folks out there who look at the results of these achievement level-settings processes and say that this is all an attempt to undermine support for American public education. And that's something that one shouldn't ignore if you're going to actually do the bigger context that Maris is proposing with respect to certainly one side of the paranoia. Because there's this other side of the paranoia too, and that has to do with, you know, what is -- why are people saying that only two percent of the kids can do something, if it isn't to draw some kind of support away from the way we have organized education in the U.S. So context here is very significant to the whole debate.

MS. RAVITCH: So in other words, Mark, please write a history of -- you have the last word, and then we're going to change places, and everybody will stretch a little.

MR. RECKASE: I don't think I could come up with a short response to that.

MS. RAVITCH: Okay, then thank you for your paper, and we'll be back in five minutes.

The Role of End-of-Course Exams and Minimum Competency Exams
in Standards-Based Reforms

Authors: John Bishop, Ferran Mane, Michael Bishop, and Joan Moriarty

Comments: Richard Murnane *and* Larry Steinberg

GENERAL DISCUSSION

MR. GRISSMER: John, you didn't put your sources in the equations, and I think that's probably important to do. We find that sources do matter across states, and that I'm not sure what the correlation between the sources and the -- but that might account for -- either make your effect stronger or weaker, I'm not sure. But I think an effort to put education resources into the equations will sort of make it a cleaner analysis. But I think I'd like to make a second point about resources. There certainly is a lot of suggestive evidence, which I think is getting stronger and stronger, that incentives are important in education. And you can get significant score gains, perhaps through that approach. There is sort of a subtle corollary with that, that I would like to shoot that in my opinion, which is that resources don't matter. That is, I think the literature is very strongly moving in the direction that resources matter mostly for at-risk and disadvantaged children. I think the new literature reviews the re-review of Hanusak's work by Alan Kruger, plus the other literature reviews that have been done, really now are coming to a strong conclusion that even the non-experimental literature says that resources matter across a number of variables. There's still a wide inconsistency in measurements, but on that it seems like resources matter. And if we really believe that resources matter for the more disadvantaged students -- and I think there's very compelling evidence for that -- most of the quality and resources is interstate, not intrastate. In other words, you could cure all of the intrastate inequities, and you still have two-thirds of the problem left, in terms of disadvantaged children getting less spent. And that's got to be a federal issue. I mean, the feds are the only ones that can really address that issue. And I think that what we'll eventually find is, is that incentives and resources work together. Resources will probably work better where there's incentives involved, but the really good news, I think is, is that incentives matter more for lower-income students, which seems like that may be the case, at least in some instances, than resources do. But it really may be possible to provide substantial increases in the test scores of disadvantaged students with both of these approaches combined.

MR. J. BISHOP: I completely agree that resources matter. In fact, in other papers, I've looked at the effects of the examination system on resources. Particularly in long-standing systems, like the Canadian systems, or the Regents, essentially the decisions that the state makes regarding how much to invest in K-12 is a response to the nature of the stakes attached to the outcomes. And if the stakes are high, i.e. your kid gets to graduate or not, based upon how good the teacher is, and how effective the teacher is, that raises the willingness of the public to pay higher salaries and to set higher standards for becoming a teacher in this system. In the New York State system, the salaries of the teachers in the secondary education system relative to the salaries of college teachers is higher in New York State than, I think, only two other states. So we are a high-paying state, generally. But we particularly like to pay our secondary school teachers more. And that, I think, is one of the effects of the fact that there's externally examined tests. Another thing that I've done in terms of looking across countries, is the relative wage of teachers is consistently higher in the countries that have externally examined systems than countries that don't. And so it creates an incentive that the public is more willing to pay their teachers well when how well the

teacher teaches really counts. And the job of a teacher in a non-externally examined system is to keep the kids in order and keep up their self-esteem and happy, and to figure out who is the smartest and, you know, to rank them. But if there's not an external assessment, the need to teach them the substance of what they're supposed to be teaching is -- often falls into the cracks. In an externally examined system that is not a solution for the teachers.

MS. RAVITCH: Ron Ferguson?

MR. FERGUSON: John, I wanted to ask you to comment on the mechanism by which you think these earning effects come about.

MR. J. BISHOP: The first effect is presumably that the students learn more, and so therefore there is a -- we of course don't know the degree to which they're --

MR. FERGUSON: But you find that in the states that have the MCE's, the scores are lower for the A students and higher -- I mean, the effect is to reduce -- scale it down for the A students and raise it for the C students.

MR. J. BISHOP: Right.

MR. FERGUSON: But the earnings effect is to raise the earnings for the A students, not the C students.

MR. J. BISHOP: Yes.

MR. FERGUSON: So it seems that if the effect was to learn more, you should have the earning effects be --

MR. J. BISHOP: Learning more effect is probably not what's driving what we're getting, because the second effect is a signaling effect. At present -- in fact, this is one area of disagreement with Larry. And that is, he was asking, is it really important if we jack up the wages, but if we basically fake the employers out into thinking that the kids are better than they really are. In fact, employers have a negative stereotype regarding high schools, which cause them to believe they're a lot worse than they really are. And that what's happening, I think, is that we're telling them that they're better than they thought, and it's more accurate. And the evidence for that is that graph I showed you about the small effect of test scores when they're very -- right out of high school, and that that's due to the ignorance of employers. And what we're tending to do then, is to reduce that weak relationship and make that relationship strong. Essentially, that relationship that Altonji found, if we were looking at minimum competency exam states or high schools, we would see the effect of the test scores or grades, in this case, on wages would be large.

MR. FERGUSON: You're telling the employers which schools the good kids come from.

MR. J. BISHOP: And that's all to the good, because it means that the employers are starting to immediately reward, rather than waiting 10 years to reward kids. That then sends a signal back to the school and to the kids for them to try harder in high school. So consequently, I see a -- the

signaling effect as a positive one, because it corrects a flaw in the labor market at present, i.e. the total ignorance that most employers have about the competencies of young people coming straight out of high school, and tends to reduce that flaw and more correctly pay them and then create the right signals back into the high school, putting kids to work. So it seems that it's working in that direction. You know, the data is consistent with that story, but you know, I still am blown away about how big those effects are, and I still -- you know, maybe when I put some more, some other variable in there, the effect will go down and become small. I don't know. We've got a lot of variables in right now.

MS. RAVITCH: Edward Fuller?

MR. FULLER: Thank you. I'd like to make two comments. One, the -- when you talked about the effect on the dynamics in the classroom between teachers and students, I actually taught at Texas under the system that I'm now studying, and it resonates a lot with my experience in that it essentially creates a common enemy for the teachers and students to gang up on, say you know, "People don't think you can do this test, I think you can, let's do it," and show them, prove them wrong. There's two problems, I think, that haven't been studied with respect to the effect on teachers and high-stakes testing and accountability, and one is that in a system like the one in Texas, it could potentially exacerbate the maldistribution of quality teachers between traditionally low-performing schools and high performing schools. And we have some data in Texas that could potentially shed light on that, though we haven't been able to look at that yet. But we know from previous studies that teachers tend to move from high-minority, high-poverty schools to more affluent, more white schools. I'm just wondering if high-stakes accountability systems might exacerbate that type of trend. The other is we hear from teachers -- my wife also taught high school, and we have a lot of friends that are teachers, and they talk about increased stress from having the additional test. Now Texas, probably in response, gave everybody a \$3,000 raise, all the teachers, but whether that's enough to overcome the stress induced by the testing, we don't know. I mean, are there higher quit rates of teachers in states that have these kinds of tests, or can you balance that out with increased pay? The second point, and I'll keep this short, is that along with what David said, accountability systems aren't really enough, and it's disconcerting when people hold up the Texas accountability system saying, "Just slap this on to your state and you'll be fine." It's what people don't understand is that there's lots of things that went into the system. One of the most important was the actual redistribution of funds from property-wealthy districts to property-unwealthy districts. And we also have a cost of education index that we're actually redoing right now, and this is for -- started in 1990, and it gives districts more money -- it gives more money to districts who have costs that they can't control due to their location or their student demographics, or those kinds of things. Essentially, the goal was to give enough money to the different districts so that they could buy the same inputs. And a study done at Texas A&M has shown that over the years of the redistribution of money, the correlation between property wealth and student performance on the test has disappeared. So I think it's really -- you can't stress it enough that that's part of the overall picture. You can't just slap on accountability.

MS. RAVITCH: Any -- Herb Walberg?

MR. WALBERG: I find, John, your analyses very convincing, especially if you look at them over many years. I also find one anomaly on them, and I wonder if you could shed any light on it. I

think that I would like to believe -- in fact, I do believe -- that externally imposed examinations have an objectivity that is extremely useful, and New York State is our outstanding example of that, as you've argued. But you've also said a few minutes ago too that resources are important, in response to David. I looked at NAEP data and spending data for the something like 40 states some years ago. New York was definitely an outlier in resource use, along with New Jersey and Connecticut. I think it was the highest in the United States. So a rival hypothesis to New York's outstanding performance would be -- aside from the external examinations, would be that they spend more money. Do you have any way to eliminate or -- that particular interpretation?

MR. J. BISHOP: We can run -- we can put it in. In previous work in an earlier paper, we did put spending, and I think the teacher -- it causes the effects of the minimum competency exam, and I guess the New York State dummy will go down when you do that. And -- but my response to David was that part of the reason why New York State spent so much money is because of the exams. So the question is, what comes first? The exam system, and that then induces more spending, in which case you want to count the increased spending as one of the effects of the exam. So not clear -- it depends on the question you want to pose. If you want to hold spending constant, what's the pure effect of an exam system? Okay, then you want to have the spending in on the right. But if you want to know what the total effect is of a policy intervention that you expect to have all sorts of other effects on decision-making, particularly if you give it time to work, and for everyone to respond. And that's why I brought up the fact that we've been very successful with getting money out of the state legislature in the last couple of years. And it's attributed by the people who engage in this battle, in part with the fact that legislators don't want to be blamed for a lot of kids failing these exams. And so it's harder for them to say no to requests for increases in funding for K-12. So --

MR. WALBERG: Although John, you'd be assuming that the entire feedback is due to the policy--

MR. J. BISHOP: Right. The entire thing about New York -- yes --

MR. WALBERG: That can't be right.

MR. J. BISHOP: Well, somewhere in between.

MR. STEINBERG: John, wouldn't it be informative to maybe do within New York State -- comparing, if you could get data on resource spending district by district, or county by county, then they would all have the Regents exams as requirements, but they might vary all over.

MR. J. BISHOP: Yes. In fact, Joan Moriarty, one of the co-authors, is working on exactly that, and she's found that spending and teacher/pupil ratios have the expected effects in -- across school districts analysis. So we are working on that as well, but it just wasn't put in this paper, though. We might try a run with the spending variable on the right to see. There's a limit to how many variables you can put in where you only have 35 or so observations. And so -- to add more variables --

MS. RAVITCH: Last comment or question from Julian Betts.

MR. BETTS: Okay, John, this is overall a very compelling paper, but you seem very surprised at the very high effects on wages. And I think I can help solve that problem for you. It seems like the story you're trying to tell here is that minimum competency exams increase skills, therefore increase productivity and wages in the labor market. But you get a very small effect on hourly wage rates, which is probably the best measure of productivity and skills, and a much bigger effect on monthly earnings. So for example, for C-minus students, hourly wages go up by, I don't know, maybe 0.2 percent. Monthly earnings go up by 6 percent. So what you're really finding here is that in states with minimum competency exams, people work more hours per month. And that sounds much less like the economist's productivity argument and more like the great training robbery argument, that education really conditions people to work harder.

MR. J. BISHOP: Maybe, or alternatively, it might -- in this labor market -- we're talking about 1992, before we got unemployment rates really down, we're talking about 1993 wage earnings. If that was just after the recession you -- I mean, part of the way productivity improvement plays out is the kids are more likely to get a job. And so -- and they work longer hours, because they're seen as more attractive by employers, because they can compete with people who have more --

MR. BETTS: But they're not producing more an hour, is the problem.

MR. J. BISHOP: --so they work longer hours. And that may be the mechanism.

Tuesday, May 16, 2000

Incentives and Equity Under Standards-Based Reform

Authors: Robert Costrell and Julian Betts

Comments: Meredith Phillips *and* Herbert Walberg

GENERAL DISCUSSION

MS. RAVITCH: Okay. Just one question I wanted to raise, which is that over the course of the last two days, several people have said, "Well, we don't want teachers to teach to the test." And even Herb said, "Well, it would be terrible if kids found out that there's going to be an essay on the Civil War," and I want to raise the objection to that, and say if teachers are not teaching to the test, and if the test is not testing what was taught, what is it testing? I mean, why teach one thing and test something else? I'm having trouble with this. I think it would be terrific if a state said, "This year, the test is going to be -- include an essay on the Civil War, and you'll be expected to answer questions about Frederick Douglas' autobiography, Lincoln's first and second inaugural, and literature will also expect you to have read MacBeth." But states don't do that. They have -- in fact, the state assessments of -- in many places are completely content-free. But I don't see a problem with teaching to the test, because I think that really raises the question of opportunity to learn. If it's not taught, then kids don't have the opportunity to learn it, and they shouldn't be tested on something that wasn't taught. And in fact, that's the way the AP course and the IB course -- the international baccalaureate course -- is organized. There's a syllabus, kids study it, kids are taught it, and then they're tested on whether or not they learned it. And I think that's much fairer than a test which hits kids with questions for which they may be, in fact, completely unprepared.

MR. WALBERG: Since I raised the reservation, Diane, let me answer that.

MS. RAVITCH: Well, there were several other people reiterated this sort of, "Well, we don't want teachers to teach to the test. I think that teaching, if it's a good test, and if it reflects what was taught, then it's much fairer than not teaching to the test."

MR. WALBERG: Well, I totally agree with you, but it --

MR. TRAIMAN: And it appears to be the legal requirement --

MR. WALBERG: -- a lot of these have to do with the nature of the implementation of it. If you said, "Oh, we're going to give you a 20-minute test, and here's the answer," some students could actually memorize the answer. So it's in how you design it, which is absolutely key. And so I think some of -- in some of these instances, it can work extremely well. In other instances, it could work very badly. In any case, it is -- it tends to be much more costly. You have to allocate more -- it has to be graded by a human being, as opposed to a machine, and people have been reluctant to put the efforts into designing excellent programs.

MS. RAVITCH: Well, I saw a presentation at a NAGB meeting last year that suggested that machines will very soon be able to do better grading of essays than human beings ever could, and

I still have trouble believing that, but it was asserted very forcefully. Richard Rothstein?

MR. ROTHSTEIN: I just want to -- this small point that you just made, there's teaching to the test and there's teaching to the test. And clearly you do not want -- if you have -- if you want -- a test can only sample a curriculum, and you want the curriculum to be broad on the particular items that are sampled on the test. So in an American history course, to take your example, you want students to know about the Civil War, but you also want them to know something about -- if you tell them in advance the only thing they'd be tested on, or what they're going to be tested on is the Civil War, then none of the other items get covered. And so it's a question of how narrowly you teach to the test. Clearly the notion that you should not show your teachers the types of things that are on the test is silly. But you can teach the test so narrowly that you're not covering anything except the particular items that are sampled on tests. And that, I think, is the problem.

MS. RAVITCH: I think that's a danger, but I think the greater danger is the kind of American history exam that I've seen from some states where kids are given a document, a so-called document, and then they're asked to read and then respond to the document. It turns out not to be a history test at all, it's a reading test. And you need bring no background information to the test, you need study no American history in order to read and react to whatever was put in your hands. Various kind of background knowledge is necessary. David Grissmer?

MR. GRISSMER: Yes, a couple of comments. This teaching to the test stuff with a couple of different perspectives. One is there's a sense that you can do something in the short term which will increase a grade, which really doesn't increase the student's achievement or wages, or -- in the long term. So clearly, what we want is methods that you can use in the short term, but you really need to look and see, make sure that whatever is taught in the short term has longer-term impacts. And they might raise test scores in grade school, you want to see higher graduation rates. And there's sort of a sense about teaching to the test, which I don't -- I haven't seen really articulated that, you know, there's some short-term fix which you could do, which will raise a grade now, but it really won't have any long-term impact. So I think long-term impact is the important part of that. The other one that I've seen is that -- that I think is legitimate -- which is teaching to the test means if you teach to this test more time, you're taking away from something else. And we really don't know what the opportunity costs are in this business yet. Whether we -- what are teachers not doing today -- when they're teaching to the test of things tested, what they used to do. And what they used to do, was that valuable, or was that just slack in the system. Are we taking up slack, or are we really beginning to substitute for something that is valuable, but not measured right now? One other comment. In terms of when we say things are expensive. You know, doing paragraph rating on tests is expensive. We're talking about, I think, even good tests cost about \$25 a student a year, and we may add \$5 a student. We spend about \$6,000 a student. So in terms of expensive, that's not a major increase. You know, we spend about probably -- what is it -- \$20 a student on R&D. What's really expensive with education is when you can't tell people how to spend the \$6,000. If we could get one percent productivity increase in the \$6,000, spending another \$5 on R&D or \$5 on a better test is a pretty trivial kind of cost.

MS. RAVITCH: Maris Vinovskis?

MR. VINOVSKIS: One of the issues that has been interesting reading the papers and listening

here is of an issue of the GED. And you mention it a little bit in your paper, but you didn't talk about it here. And I was wondering if you'd tell us a little more about it. And the reason I say that is because a lot of federal programs, for example, education for pregnant teens or young mothers who need -- push the GED, Heckman and others have now raised all sorts of questions about the value of the GED. And the question I have is, what implications are the raising of the standards for the GED? Do we do that? Does it have the same kind of tradeoffs that you mentioned there, and is there a way to add a non-cognitive component? Can you get a GED plus if you do your last semester in school, and show you can endure a semester in school? So the question is, since a lot of these students who are more near the bottom are getting GED's that are not very useful, is there a way to really work on that part of the component, or do you simply try to drive that out of the market?

MR. BETTS: I find your suggestion there very intriguing, about adding a non-cognitive rider on to the GED to have more -- a further differentiated credential, a street GED and a GED with some seat time. That would be my reaction. Of course, Dick Murnane here has done a lot of working with GED as well, which is quite interesting and quite important. So you probably want to respond, as well.

MS. RAVITCH: Dick, do you want to comment on that question?

MR. MURNANE: Yes, if I may. First of all, there will be a new GED exam in 2002. Nobody knows what it looks like. I tried to get some folks at ACE to run that, to get to work on with Lauren Resnick's new standards. They were not interested in doing that. But just on the effects of it, let me just say on the effects of the GED is we don't know what the new example will do. But I think it's -- the problem with the GED is in -- it does serve a useful role. Work that we've done said, you know, addresses the question, is it of any value to the half-a-million or so kids who drop out of school to acquire that, and many immigrants who come to this country as adults? And the answer is yes, it is of benefit, so long as these are kids who left school with very weak skills. Kids leave school for a great many reasons. Some kids leave with reasonable skills, but they often have a part-time job and the employer pulls them to ask them if they want to have a full-time job. Those kids end up earning a lot more in their mid-twenties than do kids who leave school with very weak skills. The kids who leave school with very weak skills have the problem that in they're the last hired, first fired. By the time they're 19 years old, they have resumes that have all of these blank spots. Now, what's the first thing you do when you apply for a job? Fill out an employment application in which you provide your previous employment history. And when the employer, potential employer, sees all those blank spots, you don't ever get interviewed. So the problem is for a kid who left school with very weak skills has this very spotty resume. What -- how do you indicate to potential employers that you're a different person now at age 20 than you were at age 16? And what the GED does for those kids is act as a signal A, that you will sit through a seven-and-a-half hour test, that you have reasonable cognitive skills, not great, but reasonable cognitive skills, and you have some persistence. That turns out to be worth about 15 percent in the labor market, 15 percent increase over what people would have had, which means it gets you to \$9,000 to \$11,000. In other words, not enough to -- So there is some benefit to that.

The problem is that in an environment of -- with standards, that are making it harder to get a normal high school diploma, this alternative route looks quite attractive for -- if you believe the evidence from the current population survey -- to a growing percentage of American kids. And

that's a real problem. So I don't see how you reconcile two things. Maybe Maris' idea will do this. On the one hand, having kids who made mistakes at 16 have a second chance at age 20 makes enormous sense, because it's simply the American ethic. On the other hand, providing a not very good alternative to a high school diploma at the same time you're trying to raise the stakes of a high school diploma, is deeply problematic.

MS. RAVITCH: Go ahead.

MR. BETTS: Just very briefly, I can't remember, Dick, whether it's you or Heckman and his team who reported that the average student studies something like 5 or 10 hours for the GED. But that, to me, is very important, because it suggests that really what a GED is doing is serving the sorting role, and it's not really providing the incentive role of standards. All it's really doing is identifying motivation, it's not causal. So your suggestion of linking it somehow to further course work is a very good one, I think.

MR. MURNANE: Just one more point.

MS. RAVITCH: Yes, go ahead.

MR. MURNANE: You've got to be careful with those numbers, Julian. I mean, the newest number from ACE is that the -- on average, the median among the study time is 30 hours, clearly not an enormous amount. But there's two problems there. First of all, there's an extraordinarily long right-hand tail to that distribution. So for minority people study a lot -- And secondly, the way the question is framed, "How long did you study for a GED?" And it turns out, there's a lot of people who come to that test as immigrants, spend several years learning to speak English in these ESOL programs, then a couple of years in adult education, and then finally GED preparation. And so for them, you go to them -- you know, it's a remarkable event that was also six or seven years of work. Now, the problem is does it buy you that much, unless you go on to college with it, which about 20 percent of those recipients do. That number you've got to be careful with.

MR. COSTRELL: Let me just add one more thing on the GED, because it's related not directly to this question, but to the earlier discussion of pooling and cutoffs and so on. Dick's team has done some very interesting work exploiting the fact that the cutpoints on the GED is different in different states. And so someone in a different state with the same score on the GED can fare very differently in the labor market, as I recall.

MS. RAVITCH: Alex Molnar?

MR. MOLNAR: Well, I was just intrigued by this question of productivity that Dave raised, and I wanted to get the reaction -- it seems to me that standards, at least done in the way that I personally would support, which would have them much more closely allied with the teaching and the curriculum in classrooms, that have teachers involved in this process much more directly than they currently are -- would have the short-term effect of reducing productivity. That is, you would need more teachers spending more time doing more things. The costs would go up for education at the local level. There would be a reduction in productivity. And I'm wondering how

it is that you would communicate, were we to take such an approach to standards, how would you communicate, attempt to describe the potential long-term productivity gains associated with doing that?

MR. BETTS: Mm-hmm, so really you're talking about two different things, productivity in schools, and long-term productivity in students.

MR. MOLNAR: Exactly, because I see productivity in schools going down. If there was a serious attempt to do standards the way in which I feel I could personally support them. On the other hand, one could, at least theoretically, make the argument that those productivity losses, those immediate productivity losses within schools would be made up later on by gains in the larger economy. How would you do that?

MR. COSTRELL: Well, actually I may have a different response, but what I see is evidence of quite the contrary. Just to take one example -- and this may, in part, respond also to Meredith's point about what standards imply for parents, teachers in the schools, principal agent problem kinds of issues. I have one passage in the -- we have one passage in the paper talking about a very interesting and unanticipated, perhaps, effect -- unanticipated positive effect, I would say -- of the institution of high-stakes standards, or the onset of high-stakes standards in Massachusetts, and that is the extraordinary battle taking place in Boston, as we speak, over seniority rights. The -- it's become a very public contract negotiation, quite unprecedented, in which 30 community groups, including the Urban League, the Ten Point Coalition, which is a local group of ministers and the like, have all come together to publicly support the school committee in its negotiations with the teacher's union -- it's an AFT affiliate -- to get rid of these bumping rights, so they can get fresh blood in the classroom, and so on, and they explicitly cite the stimulus of the MCAS, the high-stakes tests. So there is an incident where -- and I don't know how much it's going to generalize, I don't think -- it's going to be hard to predict, these are all going to be idiosyncratic things that take place. But clearly, it seems to be breaking down the principal agent problem a little bit, and potentially leading to a greater efficiency, greater rationalization in the allocation of resources.

MR. BETTS: I'll add an earlier paper, looking at the first year of the Chicago public schools programs, the summer school, I looked at the test score gains over the seven-week program, and estimated -- the summer school program appeared to be about seven times as cost-effective as spending during the regular school year. When I read that article I was very skeptical, because I thought a lot of this was teaching to the test, the kids would forget it immediately. There's been follow-up studies done since then which suggest that these gains, in fact, are permanent. So I'm not sure that the premise of your question is always going to be correct.

MS. RAVITCH: Susan Traiman?

MS. TRAIMAN: A comment and then a question. In the beginning of the paper, the authors define standards-based reform, and I have the sense throughout this meeting that there hasn't been a lot of clarity about what standards-based reform means, and that each of us may have a different definition. Standards assessment accountability, I think, in the view of a lot of people in the business community, isn't standards-based reform. It's the base of standards-based reform, but it's

not the reform, unless it drives the other kinds of changes that we've been talking about. Here's my question. As the business community has been encouraging employers to ask for transcripts as part of the hiring process to send signals to students that their high school records count, one of the things we've had to be very cautious about is saying, "We're talking about transcripts, high school records, not the diploma." And the reason for that is the legal concerns about the need, if you're actually going to ask for the diploma, that you show that you're talking about job-related skills. And as you talked about different kinds of diplomas, one that might be a certificate of attendance, I'm concerned that it's going to create even more complexities for employers' concerns about legal issues. And therefore, perhaps not create the kinds of incentives that you're hoping for.

MR. BETTS: Is there a lawyer in the house?

MR. COSTRELL: John, this is something you've talked about a long time, maybe you want to respond.

MR. J. BISHOP: If you've got the transcript, and the transcript's got the score on the minimum competency exam, or whatever the other -- the individual student's stakes exam is, in the case of New York State transcripts, they have each -- each Regents exam score is on the transcript. So employers may use it, if they like. In fact, hardly any are. I'm not aware of a single one that is. You've got all the information you need from the transcript, but few employers get the transcript, so -- whereas they always ask about the diploma, and presumably -- I think you're proposing the diploma thing because you're expecting employers not to get the transcripts and you want some other credential. But I think it really would be best if we created an efficient mechanism of transmitting the transcript to employers. I mean, if we can check whether a person's Visa card is valid within 10 seconds at the checkout counter, I see no reason why we can't set up a secure mechanism of transmitting a diploma -- I mean, a transcript to employers, once the student has put his pin number into the equation.

MS. RAVITCH: John, do you have your placard up? Did you have a question --

MR. J. BISHOP: Yes, I had a -- really, a comment about something that they said. You raised the -- you discussed the literature on the effects of retention, and said that the consensus is that there's a negative effect upon the individuals who are retained, but that there's deterrent effects and incentive effects maybe out from the threat of retention. I think one of the major potential reasons for the threat or deterrent effect of the possibility of retention if you don't learn during the regular school year, is to get the kids into a summer school. There's an alternative to threatening retention for states that are timid about going down the retention path. And here, remember, we're talking here about retention. This would be a judgement based partly on the tests, but partly on -- but probably predominantly on teacher judgement and other considerations that we'd make. And it's not -- would not be an absolute decision made only on the result of the cutscore, which I think would be unwise. And that is, we have a big problem getting the kids who need summer school to go to summer school. I mean, if you talk to school psychologists and other people who are involved in recommending that a particular child get extra instruction, it's often the kids who need it the most whose parents are least willing to have the kid come. Or, they say, "Okay, we'll sign up for it," but then the kid shows up only half the time. You know, there's

another instrument that might be brought about, and that is to create the opportunity for essentially the requirement to attend school for not the normal nine-month school year, but to switch to the idea of a requirement to attend school for 11 months, which we release most students from if you learn what we're expecting you to learn by the end of the school year.

MR. BETTS: Can we plagiarize that?

MR. J. BISHOP: And then the consequence --

(Laughter.)

MR. J. BISHOP: I mean, we use retention as a club to get the kid to go to summer school program. Why not make going to the summer school program a requirement, period, and then we don't have to face that tradeoff about retention. And certainly not having to go to school in the summer is a strong incentive for most kids, though it might actually be a positive incentive for many parents.

(Laughter.)

MR. J. BISHOP: But, I mean, I think we want to put -- I mean, we -- part of the this tradeoff comes from the insufficient numbers of instruments. And if we add another policy that essentially avoids one tradeoff, we -- and so we ought to have this on the table as a potential policy thing. And so I'm putting it out now, I don't think any state has thought about doing this yet, so we'll see how many states 10 years from now are doing this.

MR. COSTRELL: You will face a vested interest in the summer camp industry.

MR. J. BISHOP: Yes, and the --

MS. RAVITCH: Ed Fuller?

MR. FULLER: Thank you. First of all, that was an intellectually fascinating paper. I'd like to see you do the same type of analysis -- I think this is what Meredith was saying -- at the school level, rather than at the individual level, and I'll come back to that in a minute. On the certificate of completion, we have that in Texas, and it doesn't seem to work out quite the way that you conceptualize it in the paper. That may be because there's too small of a percentage of students that do that, but they either drop out in Texas, or they finish and successfully complete the test. I actually sat through the entire MALDEF trial court case, and one of the things I thought that equalization didn't come up in the court case because the judge said -- explicitly ruled that it's equalized -- you can't bring it into the case. I think the state would have had a harder time winning the case if funding had not been equalized. Also, teaching to the test. That opportunity to learn was also critical in that court case, in that when we put in our essential elements in 1985, they -- the state commissioned a study by James Poppum, and he surveyed all the teachers and said you know, they asked teachers, "This is what's going to be on the test. Is it fair to assume that your students will be able to answer questions on this information?" And the information the teachers said the students would not be likely to know, then that was removed from the test. It

wasn't put in as an objective. Also, they asked teachers to come in and review the test, and they looked at each individual question. They can -- and it's representative of racially, ethnically, and geographically, and the teachers can flag questions if their particular students would not understand the way that question is phrased, be it for geographic or racial, ethnic, SES reasons. But field test items, they come back and they flag the items that have a large black/white gap, or Hispanic/white gap. The teachers review all those questions again to make sure that it's fair, that all students should know that, and it's reasonable to assume that they've been taught that information. So that was a critical point in that case, also. And I'll try to finish up here. Meredith spoke about the -- you need to look at the implications for racial minority students. Initially in Texas, it looked like there might have been a slight increase in the number of students dropping out, although it's a little hard to tell, nobody's actually done a statistical analysis of it. But in the long run, it looks like it's benefited minority students. And if the court case had come about the first year of testing, the state might have lost. And it's difficult, because it's hard to put a policy in place initially that's going to work for equity. The first year or two it might look like it's working against equity. And if the court case had been at that particular point in time, the state may have lost. But with a long enough period of time, it actually showed that it did favor equity. And finally, I have a question. On your individual analysis, it looks like you kind of make the assumption that there's going to be winners and losers, and that seems to be the case when we have these tests. But on the school accountability level, I think my understanding of Colorado is that it forces a certain percentage of schools to be low-performing, or I think they give them an F, or that's the proposal. In Texas, anybody -- you could theoretically have every school at an exemplary level. And what would be the -- from an economic perspective, how do those incentives differ, or how would those different systems differ?

MR. BETTS: Let me respond to the racial gap stuff, and do you want to answer the incentives question? I think you're exactly right about this very real concern about racial gaps and what achievement standards might do. The evidence that you showed yesterday, to me at least, showed quite clearly that minorities are gaining. Chicago evidence seems to be pointing in the same direction. At the same time, I'd like to clarify our position on spending. Our point is not that school spending has a huge effect on outcomes, therefore you want to equalize. Our point is rather that we've been trying to equalize for 30 years, we're not close because -- largely because of teacher mobility and teacher preferences for where they really want to teach. That's certainly the case in California and other states that I've looked at. So our point is in states where there are court decisions about equalization or adequacy, you need to take care of that before you put standards and accountability into place, or like you said, you're just going to completely disrupt the implementation of the standards. So that's a point well taken.

MR. COSTRELL: Just to follow up on that last point, in Massachusetts, the court case in 1993, the McDuffy case, led to the -- coincided with and certainly helped frame the Massachusetts Education Reform Act of 1993, which has spent massive sums of money -- has more than doubled, went from about 1.3 to 2.8, or higher in dollars per year in local aid, raising state -- about 30 percent or 40 percent, depending on how you count it. And it's largely gone to the urban districts, to the Bostons, Lawrences, and so on, of Massachusetts. And the fact that the standards, the high-stakes standards are now only coming online after that has occurred -- has made it much more likely that the -- has strengthened the resistance to the backlash quite substantially. The urban superintendents are behind it, at least publicly, and I think privately as

well. The -- so that's a key part of my thinking, going into this. With regard to -- I'm not sure I totally understood your question about the difference between winners and loser, when there is individual accountability versus school accountability.

MR. BETTS: Yes, in Colorado, 10 percent of schools are given an F, no matter what. It doesn't seem to make much sense to do it that way, to me.

MR. COSTRELL: Well, it should be an external -- should be a --

MS. RAVITCH: Okay, since we are basically out of time, what I'd like to do is to ask the last three questioners to make the statement or question, and then if -- we'll do them one, two, three, and then if there's a response needed, you can make a response. Michael Feuer?

MR. FEUER: Well, just a quick comment then, on this business of teaching to the test. You're right, this attracts a lot of attention as to whether it's the right thing, the wrong thing, and the phrase, "All you need is the right test, and then you'd be happy to teach to it," has a lot of currency these days. But I think you need to distinguish between the use of the test for the purpose of leveraging some kind of change, versus the use of the test to get a meaningful indicator of what's going on in the school or the classroom. And so far, the evidence is, it's very hard to do both of those things simultaneously. It's not -- I wouldn't rule it out as sort of a, you know, an impossibility theorem, although sometimes I think we're close to that, but it's very important to distinguish those. So what Richard Rothstein says here about, you know, one of the problems with teaching to the test is that -- this business of sampling. But the other one, I think, is that the kind of information that you get when you have imposed those kinds of rewards or sanctions associated with a score. I mean, just to use a completely trivial analogy, if the objective was to help Americans reduce their cholesterol intake, and you had a test in which they were, you know, on a specific date they came in for the blood test and they were rewarded for showing a reduction in their cholesterol, well, you know, depending on the size of the reward, or the implicit sanction, I think you could probably get people to cut back on pizza eating for three months prior to that test, or for the two weeks prior to that test. Question is whether you have any meaningful indication, then, as to the patient's health status as a result of that, or whether you've completely undermined the quality of the information as a result of doing that. So just keep that a little bit in mind. With respect to cutscores, what we often worry about is the kids who fall below the cut and are somehow excluded or where their opportunities are thereby prevented -- where the kids are prevented from getting these opportunities. I think an almost equally compelling thing to be afraid of is about the kids who are very near the cutscore, for whom a little bit of teaching to the test will get them over the top. And then what happens is the system forgets about them. When in fact, if you look at the reliabilities of these scores around the cutpoint, well, the business of moving up or down from the cutscore is not that hard to do, folks. But the question is whether you've actually thereby, you know, completely neglected a stratum in this student population who really needs a lot of attention, but by virtue of having taught them to pass this test, they're now over the top. And finally, one question about the paying people to take the AP. I know a little bit about O'Donnell's program, it's very intriguing. I just would -- I want to ask, and maybe we don't have time for this now, but under what sort of model of American sort of democratic rule would you anticipate that sort of program going to scale? I mean, are we actually -- are you guys encouraging some kind of a nationalization of that sort of model? And if so, who is going to be

the people deciding which courses and how much, and what the reward is going to be? Just a question for the reflection.

MS. RAVITCH: We'll save it for the moment. Rob Meyer?

MR. MEYER: Yes, I wanted to talk about how -- on Meredith's point -- how student and school standards fit together. And I think there's some situations, parts of these programs where they fit together nicely, others where they're intentioned by the particular design issues, and I want to focus on one particular point. I think, you know, we've learned that as a strategic device, it may be nice, rather than just reporting their score on a NAEP test or on a state graduation test, to pick various points and label them proficient and advanced, and so on, and attach you know, a degree certification. But now as we move to the issue of evaluating schools, the fact that we've labeled various parts of the distribution, doesn't necessarily mean that the way we want to then analyze the shift of that distribution -- which is what school are trying to do is move kids up -- is to necessarily report the percent proficient. We need to ask that question, how do we want to summarize that distribution, that shift in it, where the schools that shifted up the most are doing better than the others. And we'd have preferences about whether it shifted more up at the top or the bottom. And for that, we often are going to do different things, typically, than percent proficient. And what we need to think here is we're thinking hard about the incentive effects of setting these individual standards, we need to think about the incentive effects of how we evaluate schools. And we're typically going to want to do value-added. But we know that if we set -- if we did something like percent proficient, we have this incentive for schools to allocate resources at kids right around that cutpoint, but that's not what we want. We want to do some other things, like measure the extent to which the distribution as a whole has shifted, either by looking at the mean value-added, or the mean value-added for different people, for low-achievers, middle-achievers, and high-achievers. And so I think we need to realize that different strategic devices -- that strategic devices might work for motivating individuals, does it necessarily follow us to go down a path of using those same kinds of ways of setting up the data to evaluate schools? What we need to do is value-added there, and to the extent that we can then make these systems as consistent as possible, that's a challenge to take on. But we shouldn't just go down the road of doing it one way because it seems that that's called for consistency. And unfortunately, most of the states and districts are going down the path of trying to evaluate schools by -- on the basis of percent proficient, and that's leading to distortions, the wrong incentives, and it's just not a very good summary statistic of how you evaluate schools. And you know, we know about that. But it seems like it's incredibly compelling to link the two too tightly and do it the wrong way, so --

MS. RAVITCH: Bill Honig? Last comment, and then I'll see if the respondents -- if the panelists have a comment.

MR. HONIG: Yes, I'll just keep this short. I want to make sure we don't lose Meredith's point about the difference between holding kids back at eighth grade and holding kids back at fourth grade. I have never seen a first or second grader who doesn't want to learn to read. It's not their fault they're not learning to read, it's they're not being taught. And so John's idea of a system that basically gets them the resources and puts them in summer school and doesn't put the onus on the kid is, I think, a much better way of going, and it holds the school accountable. And I wanted to ask Julian, the brief question is, it seems like in your paper you said the eighth grade

improvements were a lot greater in summer school than the fourth grade, or other improvements, which would seem to bear this out.

MR. BETTS: Right.

MR. HONIG: And the final problem is, if the -- it kind of -- it's like title one, it lets the regular school and teacher off the hook if you've got escape out of summer school, "Oh well, these kids will pick it up in summer school, it's not my problem, they're just not trying hard enough," seems to be pushing the system in the wrong direction. Did you see any -- is this resonating at all, or --

MR. BETTS: Sure, a little bit. I think that you've got a good point, and that we certainly pointed this out in the paper, that the effects in Chicago do vary across grades. And I tend to agree with you that incentives -- it's fairer to create incentives for students to work harder and have higher grades, move closer to being adults. That being said though, let's not forget that creating incentives of this sort are also indirectly creating incentives for teachers. And we can't tell in any of these exercises whether it's greater student effort, greater teacher effort, or some combination of the two that are working. In San Diego, they're just adopting a modified version of the Chicago plan which calls for summer school. And they're being a little bit more reticent about the grade retention aspect of it than Chicago is, and I think that might be a way of addressing this. On Robert's point, I think he's absolutely right that using percentage above a certain cutpoint really does threaten to -- perverse incentive to teachers to ignore all students way below or way above that point. As you've seen in our paper, we're trying to argue that you really want to have multiple measures of proficiency. One way of doing this, if you don't believe in multiple diplomas and so on, is to use mean percentile, or something like that. You can use means instead of percent above a certain cutpoint. Every single student figures into the calculation. The other thing that I think addressed Meredith's quite justified concerns about racial gaps in achievement is that you can calculate the mean performance of a school separately by races. Texas does, what California is doing is a watered-down version of the Texas program, where the API, the overall performance index, is weighted, and gives a higher weight to students with low socio-economic status. And that's a way of giving the right incentives, I think, to schools.

Academic Standards and the Social Structure of High-School Achievement
in Shaker Heights, Ohio
Author: Ronald Ferguson
Comments: Jens Ludwig *and* Wilbur Rich

GENERAL DISCUSSION

MR. J. BISHOP: Jens brought a number of things up and I just wanted to -- I think one implication of feeling that peer norms are important, and that social pressure is important, is it suggests a bias in a lot of the ways we evaluate programs. And essentially what comes out of that is the idea if you want to change things for kids, you need to change the entire school, and you need to change their entire peer group. And that if you just have a program that affects a particular kid, and then you throw them back into an environment that hasn't been changed, essentially a lot of those effects get washed out as you follow that kid through their time. They may be having affects upon the norms of the other kids, they may be benefiting the other kids. But when we do an analysis, we miss these effects. And it -- and maybe that program, the Ford Foundation thing that you -- I'd be interested in what exactly the characteristics of that program was. And that -- my impression that that was what happened in "I Have a Dream"'s experiments, too, is that -- Let me say some things about this data set we're starting to develop. We have -- we've asked them which primary school they attended, and also which middle school they intended. So there's an intention to instrument your peer group, by using where you -- what middle school you attended and what primary school you attended. So we do -- there maybe is a way to try and get around the problem you raised, of the bunch of friends you're referring to. And so that there's a lot of analysis that we'll be working on in the future, and I think Ron's made a terrific start in raising a whole bunch of provocative questions. Also, to look at just one of the schools, though the school that's given us more observations than any other school, because they did the entire -- through eighth grade.

MS. RAVITCH: Bill Galston?

MR. GALSTON: I was intrigued by the difference of emphasis among the three speakers, and coming at the same data. You know, Professor Ferguson, I think, emphasized the variable of race, more than anything else. Professor Ludwig, you know, urged that we direct our attention to issues of class, class background. And Professor Rich suggested, although he didn't quite underscore it, that maybe gender was of at least equal importance in interpreting these results. And I must say that as I just eyeballed the tables, it seemed to me that Professor Rich was really on to something. I mean, just for example, you know, I looked at tables 4, 6, and 18. And I was really struck by the differences in self-reports, not only of feelings, but also in the case of table 18, expectations about future academic performance between black males and black females. Those were dramatic differences. And I'm -- you know, this is not a literature in which I'm well-versed, but I guess my question is, A, are these sorts of gender differences, you know -- which are also visible among whites, at least in some respects pretty dramatically -- are these sorts of gender differences replicated by other data, other studies? Indeed, if they are, if this is a real difference, then how do we begin to understand it, and address it?

MR. FERGUSON: I hear -- I haven't seen it replicated, but I hear it anecdotally a lot that

resonates with this. And some of the conversations about how we deal with it that you've got to start working with black males about their perceptions and how they fit into the world. I mean, coat-switching and navigating, two words that are part of this conversation. The black kids have to understand that it's just one way you are, you can be different ways in different settings, and fit in with different people and mix. It's hard to give you an abbreviated answer. I should say one reason the paper is undone, is because there were so many gender by class, by race interactions. And trying to figure out how to organize it, when I hit on organizing the tables by level, at least then I had a systemic way for you to have a thread to go through the paper. The way -- one reason -- and I have no resistance to showing it the other way, too, but I would point out that even when you look at the black kids who have -- whose mothers and fathers both had masters degrees, that's the 18 or more years of schooling on table 2, even the majority of those kids are not in the mostly honors track. So -- and the -- there's a huge difference in which track you're in, even after you control for the family background stuff, and that's kind of what I was keying on here. But I don't have a really good answer for your question.

MS. RAVITCH: But I just have to mention -- and this is sort of glancing off this paper -- but it's been historically true that girls have always been better students than boys, and yet men control the society. And so the AAUW concluded 10 years ago that the fault lay in the schools. Girls have always done better in schools, persisted longer in schools, you know, are good students, sit in the front of the class, not in the back of the class, the boys are more unruly, therefore, it's the school's fault that men control society. It was a huge non-sequitur which led to 10 years of confused policy. Emily Wurtz?

MS. WURTZ: I'm terribly impressed by Ron's ability to, in a disciplined way, focus on a description, and nuance description that evokes reality. You kind of feel you're in the classroom, and how it is. And the question we put to you has to do with where in your own heart and head is this leading, in terms of policy. Because policy usually kills our --

(Laughter.)

MR. FERGUSON: I'm working with Shaker Heights, Ohio, I'm working with Brookline, I'm also working with a consortium of 15 schools organized by the superintendent in Illinois. And what we'll be doing is organizing variations in their professional practice, the teachers will start to experiment with and try and track what happens as they do that. And so it's -- and it's a matter of starting with some basic framing ideas, then starting to improvise a bit. I mean, I've got a long list of things we could do differently in schools to start -- to work on those things. Some stuff, it's at the policy level, but a lot of it is at the school level. The biggest thing at the policy level for this stuff is having policy not tie our hands. And in a number of these cases, we're working at the school district level, so that there's at least authority to start to fix some of these things. So I mean, we don't have time to talk about all the various kinds of things we can do. One of the things we talk about in Brookline is, just for the near term, is to put together some summer learning plans for kids to take home. The summer learning issue that Meredith talks about a lot, is have teachers work with others to suggest to students what they could do over the summer not to have that dropoff be quite what it is.

MS. RAVITCH: David Grissmer?

MR. GRISSMER: Ron's comments about misguided love, I thought were pretty perceptive. I know that Jules did an article in the class size issue that looked at, you know, what teachers do differently in large and small classes. And one of the suggestions in the article, at least, was in higher SES classes, teachers don't do much different in small and large classes, but in low SES is where you get your advantage. Because as Ron said, there's more discipline issues, administrative issues, there's less time spent on instruction in lower SES classes, and what class sizes seem to do is sort of reduce some of that confusion, chaos, gives teachers more instructional time and more individual time with students. The other sort of interesting finding in another article of Jules' was that the -- less homework gets assigned in lower SES classes than in higher SES classes. And the question is, is it because teachers -- I mean, they don't have the support at home to get their homework done, and therefore teachers don't assign it because they know it's not going to get done, which penalizes some kids. And these dynamics between both teachers and parents in this way, I think are critical. Another more interesting part to the article was the variance across teachers in how they handle these kind of things. Men teachers assign more homework than women teachers. Women teachers give more individual attention than men teachers. The variance across types of teachers of how these kinds of issues are handled is enormous. And one of the comments, this business of discouragement, I thought Meredith Phillips sort of said it well, you know, how can you distinguish between the child who puts in a huge effort and gets the same score as a child who doesn't put in the effort? And you penalize both, I mean you sort of penalize both. And what Wilbur mentioned is that messages of discouragement can be very serious for kids. And I'm reminded of -- my wife is a marriage therapist, who brought quite a wonderful set of research to my attention on how they figure out what people are going to get divorced or not. They've really done some nice research on that over longitude in the cities, and messages of discouragement were a key part of understanding the dynamics in marriages. That is, it seemed like you had to have a ratio of four or five to one positive and negative messages, and that was sort of the best predictor. I'm sure she didn't bring this to my attention entirely --

(Laughter.)

MR. GRISSMER: But I think that the level of messages of discouragement and how that can affect behavior of students is just absolutely important, down to the level of beginning to measure how many messages are getting across with a positive versus negative -- how much effect that could have in certain cases.

MS. RAVITCH: Gerunda Hughes?

MS. HUGHES: In these discussion in the last two days, I have learned a lot about the relationship between standards and culture. And one of the things I was able to appreciate is that those who set the standards do have a culture. And the extent to which children from different cultures, whether we're talking about peer cultures, school cultures, ethnic cultures, the degree to which they are able to meet the standards of the majority culture is the extent to which they are able to, as you say, navigate and work in and out of those cultures. And I think for African-American children in particular, they tend to identify with the cultures that support them most. And so in a school setting they find the most support among their peers. They do get those negative messages, and so they tend to reject that in a way that ultimately undermines their performance. So I think these discussions are good, because they help us understand what's going

on. But of course the next step is, "Okay, what are we going to do about it? Do we maintain the standards as they are, intending to do things in a one-size-fits-all, or do we really realize that there's a variability that we could exercise there, or do we adjust the ways that I measure, and then come to appreciate differences, and not of deficit models, just in terms of --"

MR. FERGUSON: I don't have a quick answer for that. It's a long conversation.

MS. HUGHES: Yes, yes it is. And there's --

MR. FERGUSON: My old view is it make some adjustments, but we don't -- we think things kids really want to know, we find ways to get them to know those things -- there's a lot -- there's a learning style that's kind of flimsy, but I think that there's just a lot we can do, not just --

MS. RAVITCH: Ron, I would like to ask you a question, which is as you looked at the behavior of kids and the influences on the peer culture, do you take any account of the effect of the counter culture, which is in many respects if not in most respects, deeply anti-academic, and deeply anti the things that schools are trying to do.

MR. FERGUSON: It's just a question about how much -- you've got one paper, or 30 pages -- I mean, Wynn Lowery is pushing me to go back 100 years and -- the story about a lot of things are what they are, and so on. So on another paper I was thinking about writing for this conference. I think I'll save it for another -- and I've already written part of it -- is one that links why test scores stop rising for black kids at the end of the 1980's, the kids in black youth culture, as part of the larger popular culture in this interactive story. The leisure reading -- the percent of black kids that report they read daily dropped to 35 percent in 1988, to 27 percent in 1992. The same here, you get a rise in -- and then that plummets by 1992. You've got major changes in behavior that -- it just happens, 1988, this happens to be the year that NTB finally gave in to have gangsta rap on TV. And then there's a whole story about what took off after that. And that's embedded in the consumerist culture which tends to, you know, lead to certain patterns of behavior and preoccupation. Just as a matter -- we talked about lifestyles. It's what the kids spend their time doing and thinking about, and how do we account for that? Some of it's supposed to be nested in their gender, racial groups, and that's nested inside a bigger societal story.

MS. HUGHES: And it's about the messages the culture sends about what's important, and how do you succeed, and how do you make it?

MR. FERGUSON: Well, it's what's important, how do you succeed, and how do you, in particular, fit into it.

MS. RAVITCH: Rob Meyer?

MR. MEYER: Yes, I wanted to comment and follow up on something that Jens mentioned. You know, the data clearly show the differences by race and also by gender as well. One of the things in the gender case is we know those -- that channel of influence is possibly direct, as the girls and boys are coming from similar families in terms of demographics. What the race affects, you know, as Jens mentioned, they could be channeling through the SES and the demographics. And

you might consider in the paper trying to, you know, use the fullest possible models to find out whether it is a direct race effect, or at least a residual race effect, after controlling for SES.

MR. FERGUSON: That's true of a race effect.

MR. MEYER: Yes, yes. And -- unless there could be something that has to do with culture, as you've been talking about. But my question is, is it -- after we would do that, is your guess that there are substantial race effects in some -- and then the following question, does it matter in terms of the policy interventions that we might eventually have at our disposal? Does it matter whether it's a substantive direct race affect, or it's just channeling through SES?

MR. FERGUSON: Yes, I'm not -- I don't care that much anymore about whether it's just -- I mean, whether it's just a race affect. I want to know what are kids doing, what are the low-achievers doing different, and what are the high-achievers worrying about that might threaten their status as high-achievers. Some of that's race, some of that's gender, some of that's social class. But the race effects, I had loaded up with all the background stuff we could find. We've got stuff about siblings, we've got whether you live with both parents, we've got those things. You load that all in, it takes up about a third of the GPA difference between black and white. Okay, what's in this paper is kind of quick and dirty, and so -- but there's still a lot of stuff left over, and I think part of what's left over is there's some pretty rich ethnographic stories about differences in the way that parents engage school. And also differences, even in Steinberg's book he talks about differences in the standards that parents set. What's the lowest grade you can take home without getting in trouble? I've gotten that data, and broken it up by race and mothers' education and so on. Asian mothers who were high school dropouts -- well, Asian kids whose mothers are high school dropouts report -- and at least of my reanalysis of the data -- about a B-minus is the lowest grade they can take home without getting in trouble. White kids whose mothers are college graduates can take home about a C-plus. Black kids whose mothers are college graduates can take home something a little bit lower than C, which is about equal to white kids whose mothers are high school graduates. So there's statistically significant differences in what kids can get away with taking home. Now, how did those standards come about? Maybe parents adapted. When we talk about these higher income districts, you know, you move in there, you think you've arrived now to the best place for your kid to go to school, and the kid comes home with C's and said, "Mommy, you don't know. I'm working hard. These kids are smart." The mother doesn't know whether to turn the screws to get him to work harder or not, so there are all these subtle dynamics. I -- my best friend in high school who got me involved in Shaker, daughter went from sixth to seventh grade, her grades dropped to D's from A's. And he asked what was going on, and it turns out she was trying to fit in with her friends and doing all this stuff, and he told her that wasn't acceptable and her grades came back up. A couple years later, she got in a class with a science teacher who didn't think she was -- could do it, and was about to push her back down into the lowest track. He went and sat in class for a week, saw this very subtle race, gender/race dynamic in the classroom. The teacher stood in the front corner of the class where the white males tended to sit, and there was kind of an old boy thing going on between them about body language and eye contact. The white girls were sitting toward the middle of the class, and the four black girls in the class were aligned along the wall, next to the window on the other side. He taught his daughter how to study science, her grades came back up, she did fine. And so there'd be subtle things going on with teachers, things that are going on with friends, all of which

are very highly race-correlated. And so you start to unbundle the race thing, you start to pick up differences in behavior that can account for why we get these -- some of these differences.

MS. RAVITCH: We've reached that point where we have to have three summary comments, and then a response from those on the panel who see the need to respond to a question. So John Hoven, why don't you begin, and then Ed Fuller, and Alex Molnar.

MR. HOVEN: Ron, do you see evidence that low-achievers behave differently in honors and non-honors classes from sampling their homework behavior?

MR. FERGUSON: No. Actually, what the black kids in the honors classes take more advantage of tutoring, they spend more time on homework. And actually the males, the black males report by a statistically insignificant margin, more homework completion than the white males do in the honors track. The white females, by a statistically significant margin, report less homework completion than the black females, less homework completion than the white males. So it's a mix. It doesn't appear to be that much of a behavioral difference. If there is a behavioral difference, it's black kids in honors when they are in class behave just as good, or better than the white kids when they are at class. The ambivalence comes in your -- what were your broader peer group outside of class, and who you identified with, and how comfortable you feel sitting in the honors class when all your friends are in the non-honors class.

MR. HOVEN: Can I take another try at that? What I meant was a particular group of students, so just taking both honors and non-honors classes. Do they alter their behavior so that --

MR. FERGUSON: Oh, depending upon whom they're around?

MR. HOVEN: Depending on -- yes, whether they're taking their honors class or whether they're in the non-honors class.

MR. FERGUSON: I don't have any data on that. I could guess, but I have no data.

MS. RAVITCH: Ed?

MR. FULLER: I just have a suggestion for Ron and John and Michael. I was interested in the misguided love suggestion that you had. And if you could have -- get data on that and look at schools in states where they start to implement standards, to see if the teachers of the -- teachers can use the standards, or the accountability system as a lever. I've heard of principals using it as a lever to get teachers to change from their misguided love. We've seen some principals in Texas do that. It doesn't -- teachers and parents and students, they weren't actually learning anything. And the principal used the standards as a lever to say, "Yes, this is great, but the kids have to read and write and do math." So if you can give that in your data a significant -- it would be really interesting to see that.

MR. FERGUSON: The superintendent, or the assistant superintendent in Brookline is trying to do that.

MR. J. BISHOP: In fact, we have data on that. I compared New York State schools, where you've got the Regents exams to Connecticut, Pennsylvania, New Jersey, and Massachusetts schools. And there was a question, if the students don't know the subject, haven't learned the material, will the teacher take it off the test? And i.e., cut back on what you expect when you think the kids are not learning. Well in New York, that was less the case than in the neighboring states. Also in New York, there was much more positive attitudes towards the teachers, the students were more likely to say they were annoyed when other students disrupted the class. The students said they were more intrinsically motivated, and there was more positive peer pressure, they said their friends were -- wanted them to do well in English and math, and so forth. So you have evidence, at least in New York versus its neighbors, of differences. But that was New York versus its neighbors. I don't think we found a difference like, between New Jersey, which has a minimum competency test, and Connecticut, that doesn't. Or -- and so consequently, I'm not so -- I'm sort of doubtful that a minimum competency test will have the cultural effects in high school that appear to be occurring for the Regents type system where there's an end-of-course exam that is part of the grade in the class, and it's embedded in the -- in what the curriculum that the students are pursuing.

MS. RAVITCH: Alex?

MR. MOLNAR: I'd just like to thank Ron for introducing a kind of flesh and blood complexity into the discussion of standards. It's been a wonderful two days for me, but the complexity of standards -- first of all, it's manifest, it's a really difficult subject to get your mind around -- but the complexity has been a kind of glass bead game complexity -- no offense to my economist colleagues --

MR. FERGUSON: I'm an economist, too.

MR. MOLNAR: Yes, I know. I know you are, but it's in the kind of bloodless complexity, which has its own elegance and beauty. What I thought of as Ron was talking and what -- the comment that he made, that you made, Ron, was that "I want to get in there, I want to do these things, and I wouldn't like policy to tie my hands." And what that got me back to considering was in these issues of standards, the questions of when it's appropriate to compete and when it's appropriate to collaborate, and how you understand which is which and when it's best, how this relates to child development, which proceeds in fits and starts and as gender in class and racial components do in many instances. And how -- to build on what Michael said -- you don't protect the value of the information you're getting from the measure that you're using to get it, so that the measure itself doesn't destroy the underlying value of what you're getting and distort your result. So you can create lots of evidence that you're making a lot of difference, but the difference has no consequence in the real world. And I think that's the really challenging part of what you had to say Ron, to me. It locates us, I think, where we always ought to be, which is back in the schools and in the classrooms as understanding the issue and the complexity of standards. So, thank you.
