THE BROOKINGS INSTITUTION

THE 2006 BROWN CENTER REPORT ON AMERICAN EDUCATION: HOW WELL ARE OUR STUDENTS LEARNING?

Wednesday, October 18, 2006
10:00 a.m. - 11:30 a.m.

The Brookings Institution
Saul/Zilkha Rooms
1775 Massachusetts Avenue, N.W.
Washington, D.C.

C O N T E N T S

* * * * *

P R O C E E D I N G S

MR. NIVOLA: Welcome to Brookings.

I am Pietro Nivola, the Director of Governance Studies which is the resource department here at Brookings under which the Brown Center on Education Policy is situated.

Today, my colleague, Tom Loveless, is going to unveil the 2006 Brown Center Report on American Education. As most of you know, I think, I am pretty sure, this is the sixth such report that the Center has published, and every edition of this report is one of its flagship signature products.

The latest edition focuses primarily on what I think are two very fascinating questions. The first is: Does the road of contentment, of happiness, if you will, as best we can measure it, of students in school make a discernible difference in their performance? The report then also takes up another somewhat unrelated question but equally important in the ongoing debate on No Child Left Behind, and that is the question of: How much truth is there to the common contention that the states are really gaming this law, artificially inflating the reported numbers of students that meet proficiency standards?

Now, the report was so assiduously embargoed, Tom, that I haven't had a chance to read it yet, but I gather that the gist of the answer to the first of these

queries is, in essence, that chummy schools really do no better than stodgy ones, or to put it another way, happiness is really no substitute for hard work at school. The answer to the second question about the No Child Left Behind Law is that contrary to conventional wisdom these days, the states are being less Machiavellian, if you will, than is generally supposed.

What is admirable about these sorts of findings is not just that they are somewhat surprising and counter-intuitive but are really based on very careful research, and that is a distinguishing feature of every one of the Brown Center's reports over the years. The research department that I head is very proud of this work, and we are really exceedingly grateful to the Brown Family Foundation for having supported it year in and year out.

With that, let me introduce my distinguished colleague, Tom, who, in addition to producing excellent work like this, is nationally, indeed internationally, recognized as one of the leading lights in the field of education policy. Among things, Tom has represented the United States at the 60-nation International Association for the Evaluation of Educational Achievement, and he is presently serving on President Bush's National Mathematics Advisory Board.

With that, Tom, it is all yours.

MR. LOVELESS: Thank you.

I left a computer on a train two weeks ago and lost it. So I am

watching my computer because I know the computer must have done it. It sprouted legs and walked away.

Thanks, Pietro, for that introduction.

As Pietro said, this is our sixth report, and what we try to do is pose what we think are interesting questions and look for data that will shed light on those questions. And so, today, I am going to talk about NAEP test, and I am going to talk about the happiness factor thing, and I am going to talk about No Child Left Behind and also whether NCLB has induced states to inflate test scores.

Let us go to the Nation's Achievement and the latest NAEP results. Now, there are two things I would like you to take away from here. These findings have been trumpeted in other analyses we have done of NAEP data. One is that there are two NAEP tests. There is a long-term trend test that has been given, the first one actually was given in 1969 in science. So we have about almost four decades now of test data.

Then there is a test known as the main NAEP which was introduced in 1990, and the main NAEP is supposed to reflect more modern, contemporary notions of mathematics. For instance, the kids are allowed to use calculators on a portion of that test. There are manipulatives that kids are allowed to use on that test. There are more open-ended items on that test.

The long-term trend is an old-fashioned test, and it doesn't change

over time, so we can measure progress over time.

We have these two tests, right, and they are both called NAEP and they confuse everybody because these results come out and they say, well, the NAEP tests say this and the NAEP tests say that, and they are two different tests.

These are the latest results from the long-term trends. The last time the results were released was 2004. This is a ballpark figure. It is not precise. If you look at the last column here, we have tried to convert the gains into changes in years of learning. So, if you look at age 9, here is the 1990 score of 230, and the last time it was given, 241. It was a gain of 11 points. That is roughly a gain of one year of learning since 1990. The age 9 kids today score about the same as an age 10 kid would have scored back in 1990.

Does that make sense? But this is all imputed and it is not terribly precise.

If you look at age 13, you see similar gains, a little over a year. Keep both of these in mind now. This is on the long-term trend, the old-fashioned test, about one year's worth of gains since 1990. This is in mathematics.

On the main NAEP, the gains are about twice as large. So we have two tests, both called NAEP. One is saying about one year's worth of gains with this youngest group in particular; the other test, this main NAEP, more contemporary, a different testing regime, is showing twice the same size of gain.

This is very interesting. We don't know exactly what is going on, and the tests, of course, test different content. This particular test, the main NAEP gives greater emphasis to things such as geometry, data analysis, problem solving. My own hunch is kids are learning more of that because more of that has been taught since 1990 than previously. On the long-term trends, those subjects don't receive as great emphasis. Things like computation and arithmetic receive more emphasis on that test, and I don't think kids are learning as much on that. We have done other studies that sort of verify this.

Here, the reading scores, the main lesson from both the NAEPs on reading is that the gains are not very large. They are not as big, certainly as the gains in math. As you see here at age 9 on the long-term trend, a little less than a year's gain. That is a decent gain. It is statistically significant, but it is less than we saw on the math test.

(Technical interruption.)

MR. LOVELESS: You see at age 13, only two months gain, so two-tenths of a year gain.

Take a look at the age 17 scores. I haven't really discussed them. With our oldest students here on the long-term reading test, we are getting actually a five-point loss from 1990. So our oldest kids on the long-term trend reading at lost ground since 1990.

Here is the main NAEP, and you can see the gains are just trivial, only two points since 1992 at both grade 4 and 8. Grade 12, the seniors in high school, we don't measure them very often. They haven't been looked at since 2002, but again you see a loss here that did occur between 1992 and 2002.

So our oldest kids, that is another big headline item here to take away today. Our oldest kids are not making much progress on either of the NAEPs. The younger kids are making the most progress. The two NAEPs differ as to how much progress is actually being made, and there are differences by subject, large gains in math, virtually no gain in reading. Those are the main headlines for the review of the NAEP data.

Let us go to the second one. This study has received a lot of attention this morning. I have been talking to people on the radio. It was fun to look at.

I just noticed for several years, looking at international data, that there is this phenomenon of high-scoring nations when kids were asked, do you enjoy math, are you good at math, in the high scoring nations, the kids tend to fall below the international average on both of those questions. They tend to say they are not all that confident in their math abilities. I am talking about both Asian and European nations. It is not confined to any one geographical area.

You have this inverse relationship between what I call the happiness factor and test scores, and the reverse is true. In the lower-scoring nations in the

world, the kids say they are happier about math and they are more confident in their abilities; fascinating.

Then I threw one more thing in here to look at with the happiness factor, confidence and joy, and then the third thing, relevance, because a lot of people today are pushing the idea of — in fact, there was just a report that came out from the video game industry — if we could combine video games and math, boy, our achievement would take off. I think these data question that.

One of the things we looked at is relevance and real world activities. In this particular database — it is from TIMSS, which is an international study — when teachers are asked whether or not they emphasize real world practicality of math, again in the countries where teachers do that, you have lower test scores; in the countries where teachers said they do not do that, you have higher test scores.

So let us look at these numbers. You have a hard copy, and you will probably have to look at it there. This is Table 2-1. I am not sure what page it is on. Maybe somebody can help me out here.

The thing to look at here is we ordered the countries by the percentage of students who agree a lot with the statement: I usually do well in mathematics. So the kids were presented with this statement, and what we are reporting is the percentage of students who say they agree a lot. There were five different categories. They could disagree a lot, and then there were in between things as well.

You can see the top countries are Jordan, Egypt, Israel, and Ghana. Here is the United States; we come in ninth. Our kids enjoy math. There is no question. We are above the international average. That is 27 percent. The U.S. is 39 percent. So they agree with that statement a lot.

Then over here are the test scores. We have put blue checks by nations that score above the international average. Now, here are the top 10 countries in enjoyment. This is at eighth grade. Here are the bottom 10 countries in enjoyment.

Take a look at Japan. Only 4 percent say they enjoy math a lot, but they score 570 as a nation which is almost 100 points above the international average. These are extremely high-scoring nations down here. Singapore is the top-scoring nation in the world, 605. Only 18 percent of Singaporean eighth graders agree with the statement that they like math a lot.

Take a look at the blue checks. They are in the high-achieving nations, and up here in the real happy countries, only two blue checks, only two countries scoring above average. So there is an opposite relationship.

Now, there is a wrinkle to the story, and I am going to tell you that in just a minute. So don't draw any conclusions yet.

I am going to show you some data within country. Everything reverses, when you look within countries. This is between countries, right, we are

comparing countries. So hang onto your conclusions.

Here is fourth grade, same pattern with the statement; I usually do well in mathematics, and percent of students agreeing a lot. There are fewer nations here, only 25. So we only listed the top five and bottom five.

Here are the most confident countries.

Here are the least confident countries: Moldova, the Netherlands, and Belgium-Flemish. All five of the countries that scored the lowest on this question had test scores above average. Only one, Cypress, was above average in the confident group. Now, that is confidence.

Here is enjoyment, same setup, The statement is made: I enjoy mathematics. Here are the top 10 nations. Kids are just having a ball in these countries as they learn math, but of the 10, 0 score above the international average of 467. In the bottom 10 countries where kids aren't quite so happy and joyful — countries like Sweden, Japan, Korea, Slovenia, the Netherlands — low numbers of kids saying they agree with that a lot, but all 10 nations score above the international average. So you are getting the picture, right? There is an inverse relationship. It is just really interesting.

It shows up at fourth grade as well. I won't belabor the point.

With the U.S., if the U.S. doesn't fall in these data, we list it at the bottom in these rankings. In fourth grade on enjoyment, in the U.S., 54 percent agree

a lot. So we just missed the cut-off here. Our kids are pretty joyful. The international average is 50. Our U.S. mean score is 518, a little bit above average as well but the same kind of inverse relationship.

We computed correlation coefficients, a statistical test of this measure, and those are in the book. I won't cover them today. They are all negative and significant, really strongly negative.

Here is the relevance question of teachers reporting students "relate what is being learned in mathematics to their daily lives," and the teachers try to do this in their classrooms. Again, an inverse relationship, the teachers who are really trying to be relevant, the nations with teachers who are really trying to be relevant, the United States scores quite high on this. This has been emphasized to teachers quite a bit over the last decade. Let us make math more practical. If kids see the practical value of mathematics, they will become more engaged with the subject is the idea. These data cast some doubt on that idea. In the U.S., we have a lot of relevant instruction going on with 66 percent of teachers in this category. Again, the high-scoring nations tend not to stress relevance; the lower-scoring nations up here are the relevant ones.

I told you there is a wrinkle, and the wrinkle is this: We have just been comparing data between countries. If you look inside any country, you get the relationship you would expect. You get the reverse. Happier kids score higher than

unhappier kids; confident kids score higher than less confident kids. You get the reverse. You don't get anything still on relevance, by the way. So relevance gets no support anywhere.

But in terms of within country data, let me show you Singapore compared to the United States, and then we will get some idea of what is going on here. I will try to explain and throw out some hypotheses. These are all correlations, so you can't make causal conclusions. I go to great lengths in the report to say I am not urging that we make kids unhappy; let us go out there and really make them unhappy. I am not saying, let us just cut them off at the knees in terms of confidence; we want really unconfident kids.

That is not what I am saying at all.

These are just correlations, but they are very interesting data, and I had not seen anyone analyze them before. So that is the idea.

Let us go back to that question about confidence, doing well in math, Singapore students and the United States. Here are the distracters, the categories of response. You can see in Singapore, 18 percent of Singaporean eighth graders agree a lot that they do well in math, and 39 percent of American eighth graders. Again, we see that American kids are much more confident in their math abilities than the kids in Singapore.

Let us compare now the scores of those categories. Here are the most

confident Singaporeans, next in terms of confidence, 616; next level in terms of confidence, 588. All of these, again, are in your booklet there, so you can follow along.

Then finally, the least confident, these are the kids who disagree a lot, 551 in Singapore. You can see the relationship I just mentioned within Singapore; the more confident kids score higher than the less confident kids.

Same thing in the U.S., here is the most confident, 541; 492 for the next level down; and then clear down here at the bottom, the least confident American eighth graders score 448.

Now, here is what I think we should take away from these data. We should be looking at this bar and this bar. If you compare the most confident American eighth grader with the least confident Singaporean, it shows the limits of self-esteem. It is not that self-esteem isn't correlated with success, but it shows you the limit. The most confident American kid still scores way below the least confident Singaporean. Just boosting confidence is not the answer to low achievement. However, within countries, it doesn't correlate with low achievement. Between countries, the correlation is negative.

What I take from that is that there are essentially national cultural settings that kids are responding to in thinking about their own success and their own abilities. Sociologists of education have studied this phenomenon for a number of

years, for about 20 years. They call it the frog pond effect. You have heard of it. If you take two frogs, two average-sized frogs and put one in a big pond and he croaks, he thinks, oh, my croak is so tiny because he is in a big pond. If you take an average-sized frog and put him in a little, tiny pond, and he croaks, he thinks, oh, I am terrific; I have the loudest croak known to a frog in frog history. Something like that could be going on here.

What sociologists have done looking at frog pond is compared kids with almost identical test scores that are in different kinds of high schools. Let us say we have two kids, Kid A and Kid B. Kid A goes to a high-achieving high school, and Kid B goes to a very low-achieving high school, but they have the same test scores.

Then you go in and you ask them questions. How good are you at math? Well, Kid A in that high-achieving high school always says, I am not very good. He looks around or she looks around, and all the other kids around them know just as much as he or she does if not more. The opposite occurs in the low-achieving schools.

This is one of the pernicious effects actually of our really low-achieving schools. Kids and parents are not getting proper and accurate signals of their own abilities. So in a low-achieving school, you may have a kid who gets an A in math, knows much more than all the other children in the classroom, but on an

international level is below average and will never know that because that student will never get the right signal as to how they are doing.

My takeaway, my practical policy lessons from this are two. One is on things such as textbooks, let us not worry so much about making them fun. There was a report yesterday, like I said, from the video game industry about really dressing up mathematics as a video game and that this would improve achievement. These data really say that is a limited pursuit. So we need to keep our eye on the ball.

Remember Singapore, the top scoring nation, the Singapore book now is being adopted in some districts in the United States. You can buy it, Singapore math. Our American textbooks are about twice as big as the Singapore text. When you dig into them and look, American textbooks in math have stories, games, lots of pretty pictures, really colorful charts and maps, and all kinds of things that are non-mathematical. They are not mathematical. Singapore texts are lean; they are in black and white; and they have math, nothing but math.

I think we need to look at our materials and benchmark them to really high-achieving countries' materials around the world; that is one suggestion. The second one — this goes back to the NAEP test — we need to make sure all of our tests because we do a lot of testing. We have a national test, the NAEP, and each of the states gives a test in No Child Left Behind. We have to make sure those tests reflect international standards, so our kids get feedback on how they are doing

compared to other kids in the world, not just Johnny and Susie in Washington, D.C. but kids all over the world. If they get more accurate signals, I think they are not going to be as happy. We might lose some points on the happiness factor, but we might gain some points on learning. So those are the recommendations.

Finally, let me turn to NCLB and state test scores. There have been about four research reports, and they were highly covered by the press. The Fordham Foundation did one study; Paul Peterson of Harvard and Frederick Hess of AEI did another one; the Education Sector released a study. They all looked at the question of our states inflating their test scores.

Here is what they did: They looked at the states' reports of the percentage of kids who are proficient — that is the term, right, proficiency — in both reading and mathematics, and then they compared that percentage to the NAEP test, the national test. That is a reasonable way to look at this. That is exactly what I would do. They found that the states are reporting about twice as many kids proficient as the NAEP test, and their conclusion was the states are inflating test scores.

Well, essentially, in this study, we did the same thing but with one little wrinkle.

By the way, their second conclusion was the reason why states are doing this is because under NCLB, the states want to look as good as possible, and so

it is giving them an incentive to inflate their test scores. That is the part that I question.

What we did is say, okay, let us look at those same data, but let us go back before NCLB and see what those gaps were, the NAEP-state gaps. What were they before NCLB because maybe the states have always done this, and that is basically, by the way, what we found. Yes, states report higher percentage of kids proficient than the NAEP test, but they did it before NCLB. So you can't blame NCLB before this particular phenomenon.

There are some other interesting things hereto in terms of these gaps. Let us take a look. This is eighth grade math. We went back to the year 2000, so it was a couple of years before President Bush signed No Child Left Behind. Here is the percentage of kids that NAEP is reporting as proficient. You can see 24.8 in 2000, 50.7 by the states, the top line is a percentage that the states are reporting as proficient in eighth grade map. You can see, yes, there is a big gap. The states definitely are reporting higher numbers, but they were doing it back here in 2000. In fact, this gap has narrowed a little bit. You would expect it to expand after No Child Left Behind if No Child Left Behind was the thing driving this inflation, right. So this gap should be widening. It is not; it is actually narrowing.

Same thing in fourth grade math, you can see the narrowing here. The gains in NAEP have been much larger than the state gains. Both are reporting gains,

but on the national exam, the gains are actually larger.

So are states reporting more proficient students? Yes, they are doing that, much more than the Federal test shows, but they always have done that and there is no evidence that NCLB is exacerbating that gap. I am going to call it a gap for a particular reason, and I will tell you about that in just a second, as opposed to this means this number is right and that number is wrong.

Here is reading. Now, in reading, you do get some evidence of a widening of this particular gap. If you want to make the case, you just stick to reading and not math. But, to me, it makes no sense why the states would say, oh gee, we will gain in the reading but not the math. That just doesn't work for me.

You can see there have been really large gains in reading at eighth grade reported by the states. The NAEP tests, as we reported earlier, just flat as a pancake when it comes to reading.

Here is fourth grade, meager gains on the Federal tests, and the states are showing much more progress. Those are the four charts in this study. Basically, our conclusion is yes, states report higher proficiency, but they have always done that.

Now, the second thing that I question with any kind of research that does this is the assumption that, well, I am a critic of the NAEP test, and I am a friendly critic. I don't have a problem with the NAEP program at all. I think it is

terrific. There are people here in the audience who work for NAEP, so I have to be sure to say that, and they do a good job. I think NAEP is a terrific program. I do have questions about the main NAEP, and I have been a critic of that test for some time.

My criticism essentially is anchored on its content validity which means does it measure what we want kids to learn. Remember the math data at fourth grade, two years of learning gained since 1990. In the Brown Center, we have conducted a number of studies that question whether there really has been a two-year gain among the younger students. We need accurate data on NAEP. I think researchers are making an assumption when they simply take NAEP and say we are going to use that to judge the states, and if the states don't produce the same numbers as NAEP, it means the states are doing something wrong. I think NAEP is not the proper test to do that with because of the content validity problem. To me, it is not testing the essential mathematics that kids need to know.

Here is one example of many, and there is quite a bit of discussion in the report about NAEP and criticisms of NAEP. This is a recent study out of the Department of Education, and it compares NAEP to TIMSS. Now, TIMSS is that international test I talked about. By the way, remember on the main NAEP, we are getting these huge gains especially in fourth grade. On TIMSS, we are getting no gain at all for the same time period. On the international assessment, it is saying that

the U.S. is making no progress or very little. Our own national assessment, the NAEP, is saying we are making such gains, frankly, and I have it in the report, I think by 2049, if the same pace continues, fourth graders will know as much as a high school senior. So we are really making gains. This is a miracle. We have put something in the water. They know so much math, but they don't seem to know so much math when they take an international test because they are getting no gains on the TIMSS.

Here is what the department did. This is a very interesting study. They took the eighth grade items on the NAEP and TIMSS, and they compared them. Both NAEP and TIMSS tested eighth grade and fourth grade, and they have frameworks that describe that content. They asked the question: If we took these eighth grade items from NAEP and reclassified them, put them on TIMSS, would they be a fourth grade item or an eighth grade item? What percentage of the items would be at a lower level on the opposite test? They did the same thing with TIMSS. They said, let us take the TIMSS eighth grade items. If we put it on NAEP, would it be reclassified as a fourth grade item?

Here is the percentage of items that would be classified at a lower grade level, at that lower grade level. On NAEP, 27 percent of the items under the category of number which have to do with arithmetic for the most part and number sense, 27 percent would be reclassified at a lower grade level on TIMSS. On TIMSS,

only 4 percent at the eighth grade would be reclassified at a lower level. In the category of measurement, it was 37-25. In geometry, a huge difference, 43 percent of the geometry items in eighth grade, on TIMSS, they wouldn't be classified as an eighth grade item. They would show up as a fourth grade item. Only 3 percent of the items on TIMSS would be so reclassified down to the lower grade level.

If you look at all these numbers and these numbers all the way across, you can see that in terms of this measure of rigor and the pitch of the test, the NAEP just doesn't look as rigorous as the TIMSS.

I am going to stop right there and leave some time for discussion. Thank you.

[Applause.]

MR. LOVELESS: Questions or comments, yes?

Do we have a microphone?

Are we recording anything? I don't think we are. Oh, we are.

QUESTIONER: That gap in the state tests and the NAEP, does that mean that the state tests are just that much easier than the NAEP or is that not safe to assume?

MR. LOVELESS: I would not assume that, and I will tell you why.

There are two things going on when you do these proficiency levels. First of all, there is the content of the test, and it could be that the state tests are just

easier and that is causing these numbers to be larger, right. There is a second issue, and that is where you draw this cut point, this line that defines proficiency. You could give a first grade level test to high school kids and draw that cut point wherever you want in terms of proficiency.

So there are two things operating here. One is the content of the tests which, by the way, none of the studies that I talked about have looked at the content of tests. As far as I know, we are the only study of this issue that has looked at the content of the tests, of the NAEP especially, and critically looked at it. The other studies just assume. Oh, it is the NAEP. It is the gold standard. That is what we use to judge states. Congress assumes that. Everyone here in D.C. assumes that. I don't. I think there are problems with NAEP.

Both California and Massachusetts, to me, have far better tests because they are more content-driven. They have really solid mathematics.

We had a panel of mathematicians in November here at Brookings, and I gave them the NAEP public release items. NAEP, when they are done using items and they are not going to be used again, they put them on the web and release them. We took all the items that were coded algebra. We had three mathematicians come in, world class mathematicians — one from Stanford, one from Yale, and one from Michigan — and asked them a simple question: Is this algebra? And they said no, this isn't algebra; you can't call this algebra. On at least 25 percent of the items,

they felt that they were algebra items. They were mathematically algebra.

We are gaining in math on our national test on how well our kids are learning algebra in the eighth grade with a test that mathematicians have some questions about. That is again another issue about the content validity of NAEP.

But there are two things going on. One is the content of the tests. The second one is where do you draw these cut points.

QUESTIONER: As you know, the Department of Education and even President Bush have said No Child Left Behind is working because test scores are up, and they usually cite nine-year-old NAEP scores. What is your take on their claim?

MR. LOVELESS: I think it is a very weak claim. I think it is a very weak claim.

No Child Left Behind was signed in 2002 and began to be implemented in 2003. All the provisions of No Child Left Behind haven't really been fully implemented yet, pretty much, but it is hard to pinpoint one wiggle in test scores and say No Child Left Behind caused it. I am not convinced at all.

QUESTIONER: I would like to know your take on people often drawing comparisons to the United States and what we know in comparison to Singapore, which has the population of Chicago and tends to be less diverse socially and culturally than the United States, and saying let us draw curriculum from a country like that and trying to stick it in a country where it is totally different.

MR. LOVELESS: Right; I think that is a good point. Each of these countries is different culturally, but I suspect the demographic differences are pretty much irrelevant. So I think the diversity question is probably pretty much irrelevant.

But I do think you are hitting on something, and that is the cultural differences. There are some important cultural differences. One of them is, for instance, remember on NAEP, our oldest kids, our high school kids do really poorly in terms of progress. They have not made much progress. When you look at how education is thought of as part of a teenager's life in Europe and Asia, it is totally different than the United States.

We know this on such basic things as how do our teenagers spend their time outside of school. Forget about school, outside of school. We know that we go to school less. We have a shorter school year. We have a shorter school day, and we don't do as much homework. Even if you look outside of school and you ask the question, how do kids spend their time? American teens are dominated by three big activities. One is hanging out with their friends, and I am not against that either. The second one is sports and extracurricular activities. The third one is working part-time, part-time jobs.

Now, let us just take that third one. Over half of American seniors work part-time at some point during the school year. Around the world, this is unheard of. It is absolutely unheard of. In fact, it is a stigma in most of Europe and

Asia if you work when you are a teenager, it means something bad. It means your parents don't care enough about education. Your family needs money. In the U.S., it is thought of as now you can buy CDs and DVDs and stuff like that, and you learn work skills and other things. There is no evidence of that at all.

My point is this: Those are all cultural differences that are costing us in terms of academic achievement. In Tom Friedman's book on the flat world thing, I have gone blank on the name of the book but anyway the flattening of the world economically. It is not flat. What is it?

QUESTIONER: *The World Is Flat*.

MR. LOVELESS: Okay, I am close. I am getting close.

In Friedman's book, remember he visits India, and the one thing he comes away with in terms of education, he says, look, our teenagers who are in the mall, chatting on their cell phones, they are competing with Indian teenagers who are working their tails off to learn calculus at age 16. So, again, that ties in with the happiness factor thing, not that we need to turn American teens or even attempt to because there will be a revolution if we tried. I taught school a long time, so I know what teenagers are like. We can't turn them into Koreans or Singaporeans or Belgians. We just can't do that.

But what we can do is make it very clear to them that their main responsibility is not work at McDonald's and be closing on a school night at

midnight, but their main responsibility is to do well in school and to learn as much as they can, and we don't do that as a culture. We just don't do it.

QUESTIONER: Hi. I would like to get your take on the textbooks. You mentioned that it is a lot of pictures and stories, and it is very fat, and the shift in curriculum has gone to standardized testing especially the mathematics. What is your take on the reasoning or what the timeline is in this shift from standard mathematics teaching to ridiculously long story problems and fluff like that?

MR. LOVELESS: Well, it was accelerated in the late eighties, early nineties in terms of the shift towards what I would refer to as non-mathematical material coming into math books. The textbook publishers, to be fair to them, first of all, they are in the business of making money. They are selling to customers who are demanding certain things. They are giving what American schools and parents want, which is big, fat, fluffy books that are enjoyable to work with. They market these things very well, and they do focus groups. They are not fools. they know what they are doing. So they are giving their customers what their customers are asking for.

Also, to be fair to the publishers, and I don't want to be too fair but to be fair, they say, look, we have 50 states with 50 different standards. We need to write books that meet all 50 states' standards, and this inflates the size of the books. That is their reasoning.

I think the real key here is we have essentially arrived at the

conclusion that American kids can't learn math as math, that it needs to be dressed up as something else, and I think that is a mistake.

QUESTIONER: I was wondering if the report addresses teaching pedagogies and what your take is on the relevance or importance of that.

MR. LOVELESS: We don't. We don't in this report. We have looked at pedagogy and different instructional strategies before. My own take on it is that we did adopt a number of instructional approaches.

Again, a lot of this was in response to the National Council of Teachers in Mathematics and their standards document of 1989 which was very influential, and it urged a number of practices that I think just haven't borne out in terms of higher achievement. Working with calculators with very young children, the NCTM recommended in 1989 that beginning in kindergarten that kids work with calculators — there is no clear evidence that is either positive or negative on achievement, but I have my own questions about it as an old elementary school teacher — and working in groups all of the time instead of direct instruction by teachers.

We do know in reading, we have very good research on what teachers need to be doing in reading to get kids to learn, and it has to do with, for the most, part a very directed instructional role for the teacher, not student-centered, not always group learning. Cooperative learning can be very effective if it is structured correctly

by the teacher, but it requires the teacher's input. So we know a lot about reading.

To be honest with you, the research base on mathematics just isn't as robust as it is in reading. We can't give a whole lot. We can give some guidance but not like in reading in terms of telling teachers, if you do this, it will pay off in higher achievement. We have a lot of work to do in terms of research on what are the most effective ways of teaching math.

QUESTIONER: I was a member of NAEP at the time the proficiency standards —

MR. LOVELESS: NAGB, yes.

QUESTIONER: — NAGB — were accepted. I was the only person to vote against it.

MR. LOVELESS: That is terrific.

QUESTIONER: I am delighted, therefore, with your questioning the content validity.

Secondly, I understand the frog pond issue, but I also know that in Asia particularly, the tendency is for people to react shyly in terms of a question about confidence. Did you take that into consideration about their own assessments of themselves?

MR. LOVELESS: Yes; you are right, and that could explain the Asian nations, but there are a lot of European nations on that list too and it doesn't

really explain. There are enough nations with this pattern that those kinds of differences should wash out. In other words, there are other cultures where the kids maybe aren't so shy to talk about their confidence, and yet you are getting the same effect. And vice-versa, there are countries on the other end that are low-scoring with the opposite. That could be part of the explanation, but it certainly isn't all of it.

QUESTIONER: You talked about the no correlation or even a negative correlation between relevance and learning. I am wondering how you can actually test relevance on a standardized test.

MR. LOVELESS: You can't, and so the measure here is a survey of teachers, just asking them whether they stressed the daily practicality of mathematics and the question was in over half their lessons. It is trying to isolate those teachers who really did think relevance was important. But you are right; that is a dirty measure of relevance.

QUESTIONER: But how do you even test with the standardized test? How do you use relevant questions? How do you get that the kids are learning? In other words, if you are teaching them relevant things in their daily lives, you can't necessarily test whether or not they have learned that even.

MR. LOVELESS: I can give you an example of what I mean by a relevant exercise. There was a book called *Mathland* that I was particularly critical of in the 1990s. It was the number one book in California, used in elementary

schools. It didn't have a book. It wasn't a book. It was a kit. It was a little box of activities for K through sixth grade. The kids had various what were called explorations that lasted about two weeks.

Now, the one in particular that I thought was nonsensical was teaching bar graphs. Nothing wrong with teaching how to do bar graphs, but you can honestly instruct kids on a bar graph and how they work and do it in an engaging way in about a half-hour. This was a two-week exercise where kids first polled each other on their favorite rap musicians, and then they went around the neighborhood and polled their neighbors on various things. And they built these meticulous bar graphs, again always coming together in groups and discussing their bar graphs.

Well, that is just a waste of time. I mean to take two weeks to do something that the Singapore math text does in a half-page, again, it is inefficient and it is not going to lead to higher achievement if we spend that much time doing something that is trivial. That is what I mean by relevant.

The authors of these books, because they are old people like me, right, half the time they don't even know what is relevant to a little kid, to a fourth grader. So it also gets patronizing. It is usually about music or sports or something, and half the time the questions are asked in such a way that no sentient eighth grader would even ask it that way. You can tell it is an older person posing the question.

We have to be very careful of this whole idea of relevance. There is a

liberal tradition in the world that says these are disciplines. Math is a discipline, and some really great people who are brilliant discovered various things in math. Let us just teach that. That is good enough all by itself. It doesn't need to be fluffed up and made into cotton candy.

QUESTIONER: As you know, Tom, a lot of people are coming to the conclusion that as we go up the age scale, we are doing worse and worse relative to international comparisons and that the real focus of attention in coming years ought to be on high schools more than it has been. There is nothing in this report to disabuse anyone of that notion. Do you agree and, if so, what is your interpretation of this pattern?

MR. LOVELESS: I do agree, and I think the NAEP data are pretty clear on that and the TIMSS data are all very clear on that. This is not a case where all tests are saying something different. They really are saying that we not making gains, sufficient gains, at the high school level or we are very stagnant there and not making progress.

I don't know what the solution is. I do know that the solutions that are out there now don't appeal to me very much. I don't think they are backed with very much evidence. Instead of just saying let us try some different things and we really don't know how to fix high school, I think we have people rushing in with billions and billions of dollars — I am thinking now of foundations — on some dubious

pursuits that I don't know are going to pay off, but I don't know the real answer.

The left, the liberals typically argue for more investments in education, more money. Conservatives argue for choice, more vouchers, charter schools, this kind of thing. Both of those things move the bar, but they move it about that much. They are both positive, but they just don't have the kind of impact that we would want to fix the problem.

QUESTIONER: Thank you very much for your presentation.

MR. LOVELESS: Thank you.

QUESTIONER: You mentioned that you ran a number of correlations, not necessarily taking into account causal effects. I am wondering if you could speak more on these causal effects. I am particularly interested in if this includes methodological strategies in your sample population, for example, regional differences within a country, and also social stratification. Thank you.

MR. LOVELESS: We didn't control for any of those things, statistically. In terms of the correlations, they strictly were just correlations of two variables without any control, so there was no multi-variant regression or anything like that. In these data, even if you ran that kind of sophisticated analysis and tried controlling for other variables, that doesn't take away from the fact that in the end, these are just correlations. In the end, these are just cross-sectional data of how do nations achieve at one point in time. We don't have longitudinal data where we

could control for initial differences and then map growth in an accurate way.

You just can't push these data too far. It is just better to say, I think: We took a snapshot of achievement around the world. Here is what it is saying about these two phenomena. I think it is interesting. It at least suggests that we need to be taking a hard look at these questions.

But you cannot infer causality from these data, nor from NAEP data. Same situation, you can't infer any causality there either.

QUESTIONER: Education Weekly, I don't know how long ago, published an article entitled "Johnny Can't Read In South Carolina, But If He Moves To Texas, He Flies With The Stars." Isn't that evident that perhaps tests in general just aren't a decent way of taking a look at what is going on inside the American classroom?

For example, the reason why high schools students are performing not where they should be is because they are not reading books anymore. They are too focused on making PowerPoints and something else. Spending time in a Japanese classroom for four months, you don't see computers anywhere, and this is Japan, a highly advanced technology country. Is there perhaps maybe a prompt to move away from just looking at tests and find another way to measure performance in a classroom or why we are where we are?

MR. LOVELESS: Yes, well, the whole testing regime in the United

States is fairly new. It really is a fairly new idea that we would measure, regularly measure what kids learn. When I was growing up — I grew up in California — we did have a state testing program, and we were tested but not every year. We were tested, I think, every three years on a test that was known as a CAP test.

This idea that, look, we say to schools, here is what you should be teaching and then we go in and measure whether or not the schools are actually teaching it and whether the kids have learned it, that is a relatively new idea in the United States. It is not new around the world. It is not new in some parts of the United States. Of course, the New York Regents test is over 100 years old. So we have done it before.

As to the question of should we be doing something else, I doubt it. I mean as bad as tests are, the alternatives strike me as even worse. For example, grades, grades are just notoriously poor indicators to compare Kid A to Kid B when you are in different schools and different classrooms. Teachers have different grading criteria. Those just aren't very good indicators.

Remember before, I talked about the frog pond effect. There have been studies looking at kids with similar test scores in extremely high poverty schools and comparing them to wealthy schools in suburbs. The kids with As in the high poverty schools would have an average grade of a C out in the wealthy suburbs. Again, these alternative signals to give to kids could be really worse and part of a

bigger problem.

QUESTIONER: Thank you for your fine presentation.

If you were a researcher and were interested in comparing how kids are doing over time and across the states and you mentioned that you have some issues with NAEP data, which kind of tests or what would you look at? Also, I am particularly interested in the limited English proficient students. NAEP has some education of how the kids are doing, but after your presentation, I am not sure if that is the right way to look at things.

MR. LOVELESS: I will speak to math in particular because that is what I have done the most work with. In math, what I would change about the NAEP at fourth grade, I think a fourth grader must know arithmetic, a fourth grader. So I would be measuring whole number arithmetic — addition, subtraction, multiplication, and division; some very light knowledge of fractions and decimals, they might have some idea.

I would not be measuring geometry because it turns out to be things like do you know the shapes of squares and rectangles and circles. Is this a circle or a rectangle? Who cares? I think that is nothing to build on at fourth grade. So I would eliminate geometry.

I would eliminate data analysis. I really don't care if fourth graders can tell you something about a bar graph or a pie graph. I don't think that is

necessary, and they can learn that later.

I would focus on what is happening around the world. Most countries don't measure all the things that we measure. Let me put it to you this way. Imagine you meet a ten-year-old who comes out of fourth grade, and the student knows probability and statistics, knows what graphs are, knows geometry, knows measurement and various measurement units but can't add or subtract or multiply or divide. Now, I am sorry; that student does not know mathematics and is no position to go on to learn things like algebra. So I think we need to really focus our tests at the fourth grade. At the eighth grade, I would obviously be doing some other things, and I would be broadening it out to include geometry and algebra.

The algebra items on the fourth grade test to me are not algebra. That, in particular, is the place on NAEP where the items do not indicate the content of the item that is supposedly being assessed, that the item is assessing. There are algebra items on NAEP that don't have numbers. They have letters. They have like ABC, AB blank. What letter comes next? These are pattern problems. We do not know from research that that is correlated with later success in algebra, and I think we should stop assessing it. That is the kind of thing I would get rid of.

QUESTIONER: (off mike) If we don't look at the NAEP, where would we go without the NAEP?

MR. LOVELESS: If we don't look at NAEP, like I said, California

and Massachusetts, I think, both have excellent math tests, and I would rely on their content validity. I think the TIMSS test is a better test than NAEP because it focuses on the content that I just mentioned as opposed to the less important stuff.

QUESTIONER: You talked about relevance of culture and curriculum but kind of the irrelevance of diversity. In testing to measure performance, what about diversity in that testing?

MR. LOVELESS: Obviously, we need to measure that, and we do. It is very important to look, within countries especially, at demographic differences, test score gaps. I didn't mean to imply from my question earlier that we should not be doing that. We should be doing that. We should be concerned about it because the test score gaps in the United States are huge, and they indicate problems of equality that need to be addressed.

What I was referring to was with all the questions after American teens are getting the same sorts of cultural cues that math achievement is unimportant as so where my white jeans are. So I am just talking about larger issues that are affecting all demographic groups. That is what I meant to suggest, but thank you for clarifying that.

QUESTIONER: I guess as part of the whole self-esteem culture in the United States, we are not holding back as many kids as we probably could potentially be doing, but I guess from this data, it shows that they might not even potentially

learn anything anyway if we did hold them back. How could we hold kids back effectively in your opinion?

MR. LOVELESS: The issue of social promotion and retaining kids is very controversial. Here is where I will side with the warm caring people. Actually, the evidence to me is fairly persuasive to me that retaining kids does nothing for them and probably even harms achievement. If you look at carefully controlled studies and ask the question, do kids that go on score better or worse than kids who are retained, it turns out that they score better even though they didn't know the material they were supposed to have known.

Now, some of this research is dated and it goes back to the sixties and seventies when there were a lot of these studies done. Frankly, as a nation, we were retaining kids for non-academic reasons a lot of times. I know in my elementary, occasionally, you would hear a teacher who would have a conference as a faculty on retainees, and we would go over the kids who were being held back. Sometimes you would get teachers who would say things like, well, he still has his baby teeth. Well, who cares? I may still have some baby teeth. I don't know.

This idea of maturity and using these non-academic reasons, I think we don't do that very much anymore. A lot of that research has to be updated. The bottom line here is my reading of the retention research is it does no benefit to retain a kid, and I think it could actually be harmful.

QUESTIONER: In your study, you question, you suggest that states should not model their math tests after NAEP, right, because the NAEP is deficient because of content. Should states look at the TIMSS then? Should they look to that test as a better test that they should move in the direction of modeling tests after TIMSS?

MR. LOVELESS: Yes, I would like states and the Federal Government to look more to international world standards of what is important in math to know rather than the current frameworks that we have. That is right. To me, TIMSS would be an improvement. TIMSS isn't perfect either, but it would be an improvement.

QUESTIONER: When you talked about the activity sort of book in California, has that approach been repudiated in California? Is that book still around?

MR. LOVELESS: Oh, *Mathland*, the book fell out of favor. It probably is still used, but it is not the number one book anymore.

QUESTIONER: In general, do you see math instruction going back to a little more of the basics that you described?

MR. LOVELESS: Yes, I do. There have been actually encouraging developments on that front. The National Council of Teachers of Mathematics that I mentioned put out, released a document three or four weeks ago called *Curriculum*

Focal Points. Take a look at it. It is excellent. For the first time in my life, I praised NCTM for this wonderful document. It really is good, and it lays out in a clear concise manner. First of all, it is only like 40-something pages long. The old standards were hundreds and hundreds of pages long, and they were filled with fuzzy stuff. This lays out what kids should know and when they should know, and it is very heavy on what I mentioned which has to do with number sense, arithmetic, how to manipulate numbers. We want young children to be doing this in particular, so that it sets the stage for them to go on to more abstract math.

I think the trend is actually towards a leaner, meaner mathematics in the United States. With NCTM, I hope they just really keep at it. They are now headed in the right direction. I wish they had done this 20 years ago, but I am really pleased with that.

QUESTIONER: You analyzed the NAEP math test and found it deficient. I wonder if you have looked at the NAEP reading test and what your analysis of that is and whether that is adequate or how that might be changed.

MR. LOVELESS: We are going to do some work with the reading test probably next year, but I haven't done it rigorously. I have looked at items.

The other one that interests me is the science test, and I have not looked at that either in any kind of rigorous fashion. We are getting different results on the science test, too. In the main and the long-term trend, especially with the

oldest group of kids, twelfth graders and age 17, one test is saying they are making gains, and the other one is saying they are losing, that their achievement is declining. So I don't know what is going on there. I would like to look at both of those.

Other questions?

QUESTIONER: It is not a question, just a comment. The Nobel Prize Winners, the sweep that the U.S. has just had, is it the old math? How do we account for that sweep?

Also, in terms of what Asia still seems to want from it, and I speak to that because I have done some work in Asia, creativity, innovation. Over there, they are very envious of what we have been able to do there.

But go back and talk about the Nobel Prize Winners first.

MR. LOVELESS: I have no idea what kind of math instruction the Nobel Prize Winners had, but I remember Mort Sahl. I remember Mort Sahl, the old comedian, used to talk about the first something like 12 Presidents of the United States didn't go to schools because there weren't common schools. So he drew on that. He said, but our last 12 have, and if you look at the relative quality, maybe we should just abolish schooling altogether. In fact, he called it counter-Darwinian or evolution in terms of Presidents.

To the second question, you are quite right. If you tour Asian nations, and I have done some visiting in Asian nations including I took a group of high

school kids on a tour of Asian nations for six weeks in 1985. It was fascinating. We stayed on a communal farm in China and went to Taiwan and Japan and Hong Kong. You are quite right. All the adults we came in contact with just admired these kids. Look how independent they are, how creative they are, how confident they are, and how joyful they are, everything we see. Boy, they don't know math, but that is all right. You are right.

The Pew Foundation did a study of parents, and this was just about a month ago. It was fascinating. It shows you these kinds of cultural differences. They asked parents: What do you really worry about? Well, the parents in Japan, China, and India are worried that their kids are under too much stress. Maybe we are just pushing them a little too hard with schooling. We are not letting them have enough fun.

In the United States, it was the opposite. Parents were saying, we don't push our kids enough. We don't teach them enough. We don't make them work hard enough.

So parents are probably getting a pretty accurate idea of where the extremes are in our two cultures in terms of these different cultures. I didn't answer any of your question, but.

QUESTIONER: (off mike) They were more comments.

MR. LOVELESS: Okay, thanks.

Any other questions?

Well, thank you all for coming, and I will talk to you again.

(Applause.)

* * * * *