

The States' Impact on Federal Education Policy Project: New Perspectives on Research and the Historical Record

Transcript of Session 2: The Rise of State Standards, Assessments, and Accountability

Lynn Olson: Thank you. So we got off to a great start this morning talking about this whole issue of federalism. I just have to say when Gordon asked me to serve on the adviser group I did so because the issue, in some ways, is near and dear to my heart. I have had this incredible privilege and honor, really, of spending twenty-four years at *Education Week* chronicling state efforts to improve education and state interactions – tense, more tense, less so tense – with the federal government about how that improvement effort should occur. So I also have a strong and abiding interest, having spent twenty-four years trying to chronicle this movement, in making the work of states transparent and accessible to reporters, researchers and others. And really, as public agencies, consider that part of the public duty of states is to make as much of this accessible as possible. So I am thrilled to be here and really looking forward to the conversation.

Rather than do lengthy introductions I think we will give Mike some time to talk. Let me just say that we really have a great panel here. I don't think you need to introduce Mike Smith other than to say the paper that he wrote with Jennifer O'Day on systemic reform, I really think provided some of the framing intellectual concepts for the movement we now think of as standards-based education. So it is great to have him reflect on what has happened since. Then we have Marguerite Clarke who can talk to us both about the state impact of some of those movements, and also put it in a global perspective. How do other countries approach this whole issue of school-based improvement and standards-based reform? Finally Carmen, who can talk about this notion that much of the federal activity has been focused on students with special needs, or students who have not been well-served traditionally by states and localities, and what has been the impact for English-language learners and limited-English proficient students, of activity in this area both at the federal and state level, and how did they interact. So with that I am going to turn it over to Mike.

Mike Smith: Thank you, Lynn. Before I get started with my talk, I would like to ask us all to give Lynn a strong round of applause. This is her last day at *Education Week*. She has been there for twenty-four years and she has been clever enough to negotiate a move to the west coast where she can get a bigger view of what is happening in states and the federal government. She will be working with the Gates Foundation. So if you will just join me. She has written probably the best stuff, undoubtedly, from a reporter's side, but also she has gone much deeper than the typical reporter to do really serious analysis of what the issues are and how they have changed over time, and been one of the leading people in the yearly review of how are we doing as a country and so on. So it's a great loss and for the Gates Foundation it's a terrific gain.

Second, a comment I have is that I thought, as Lynn said, I thought the first session was great. It framed a bunch of issues in terms of the way political scientists think, which gives us a real understanding, I think, of some of the complexity of what we are going to be talking about. I am

not going to try to frame it in those same terms, but as you listen to me, be thinking about the notions of fragmented centralism and its various forms. Be thinking about the venue shopping, and definitely be thinking about the implementation issues and transparency. These are all issues that come up and are of tremendous importance as we think about the nature of policy and education policy in this country.

Just as a very quick capture, we tend in the United States, at both the federal and state level, with respect to the implementation issues, to be very much like some of our corporations. It's those corporations that only look to the next quarter. We tend to think that we need something done immediately, even though getting it done is a very complex thing to do. So that's a kind of a theme throughout my entire discussion today – that you have this implementation problem that exists, and somehow we have got to come to grips with it.

So I will go through this. It's a little bit different than the version you've got in your handout. There is a couple of the slides taken out and a couple more slides put in, so this will be a little bit different but you can generally follow it along from what you did before. I am not going to go into this in any detail in the overview because we are going to do it right now, so you don't need it.

Short history, and this is a very short history. If you track back and you look at some of the work that Maris Vinovskis, who is here somewhere, has done, he did a little piece on systemic reform. In his analysis one of the things, he tracked back the set of ideas around standards-based reform and so on back to the notions of coherence in the school and thinking about taking that coherence from the school to the district to the state, and possibly to the federal level. So the notion of coherence is a second notion that is going to be dominant here.

Standards themselves as content standards, what children should know and be able to do, became kind of the catch word around Washington and a lot of policy circles in the '80s, particularly in the late '80s. In part, what this was, was simply stealing. This was stealing from some other countries in terms of ideas. It was stealing from some of the states that are mentioned up here. I mean California was evolving into...it had something called curriculum frameworks which are very similar to the standards that we think about now. Kentucky went through an entire, huge reform around standards in the late '80s and early '90s when its court actually threw out the entire system. So they were able to take a really big step in Kentucky and change the whole system around, a much more coherent approach to the problems.

The first Bush administration was on top of this and Diane Ravitch was a real advocate of standards. The first way that the term came in was really in the NCTM standards. The National Council of Teachers of Mathematics put together this set of standards over a three or four year period of time that really began to shape the text books and so on. So its work was continually referred to as people got into this set of issues. In George Bush one's, in his group both the Department of Education and the National Science Foundation and the Humanities Institution created standards and put them out as voluntary standards.

So we have three kinds of standards that we need to talk about, and it is really important to think of these. Content standards, these are general specifications, or they could be specific

specifications, of what children should know and be able to do. The way they were initially envisioned was as general specifications but they have evolved, I think, a little bit into much more specific.

Performance standards have now evolved into cutoff or cut points on standardized norm reference tests. What they were initially envisioned as were actual performances of students who could do a certain kind of academic task. So the problem is this, that if we say in a content standard that a child should be able to write a short paragraph on a certain book...on a book. So this is the standard, let's say, for second grade. Write a short paragraph on a book. What is really meant by that? The idea behind performance standards was to begin to spell out what was meant by that.

The Australians put together a set of performance standards that were really quite incredible. They'd actually pick books. They'd get students to write about them. They'd take copies of that written work and they would rate that written work to show what is really good work, what is mediocre work, what is poor work. They would make that transparent and available to teachers around the country, and in doing so what they are doing is communicating what is the serious... what is the real expectation here? The expectation is not just to write any paragraph. It's to write this kind of paragraph – the paragraph that had this kind of sentence structure, this kind of argument, this kind of logic and so on.

You can take that kind of argument and apply it to any of the content areas. That is what a performance standard should really be. It should be a reflection of student work. We used to think about having one of these cocktail table books. People don't have cocktails at home anymore but they drink wine, so it is a wine table book. It would have in it student work from all over the world at different grade levels or at different ages. How well do they write? You could compare that. People could compare that to what their students were doing in their own school and begin to get a real sense about what kinds of work their students were doing and what is being asked of them in their school.

The final kind of standard here is an “opportunity to learn” standard. This is something we don't talk about much anymore. Bill Taylor and I were talking about it last night in kind of a dreamy way. An “opportunity to learn” standard is simply the idea that if you are going to hold somebody to a high standard, you ought to give them the chance to actually get to that standard. What is it that they need to get to that standard? So that led to the idea of “opportunity to learn” standards. There are a lot of us who believe that accountability really can't be carried out unless you have the opportunity to meet the level that you are expected to reach. So fair accountability means you have the opportunity to learn to that level. It doesn't mean that everybody is going to learn to that level, but it does mean that everybody gets a chance to. So that is what the opportunity to learn standards are.

What motivated state systemic reform? There is a vision, of course, that standards... you hear often that, well, there's standards and there's accountability. That is what we have in the United States. We have a standards and accountability system. But there is something else in the middle and that comes back to the word “coherence” that I was talking about a few minutes ago. The observation in the late '90s, and even before that, was that federal and many state policies are

incoherent. You heard a lot about that this morning in the first session; kind of a cacophony of discordance when you put together the federal level, the state level, the local level and so on. So that was one concern.

Second concern was that there were two curriculums, and there still are in the United States in many places. There is a curriculum for the well-to-do and there is a curriculum for the poor. The fundamental concept here was that the curriculum for the poor was a watered down version of curriculum for the well-to-do, and therefore the kids who got this curriculum, received this curriculum, would never catch up. If they don't get the material they are not going to learn the material. They don't get algebra; they are not going to learn algebra etc., etc. You can just play it right out so if have a different curriculum then you are doomed to failure.

Finally the idea that organizational theory, I mean of other kinds of businesses, private businesses, but also hospitals and other places, had a body of literature which suggested that if you begin to align your resources you are going to be more effective. You need to align them around something. What you aligned them around was standards. Those became your goals, in effect, content standards and performance standards. You then align resources around those and that became the driver in thinking about changing the system.

Why the states? That is an issue I think that came up again a lot this morning. Lots of reasons for it. The tenth amendment is one reason, obviously. These layers of policy is another reason. The way I like to think about this is to divide policies into two categories. Now it's very simplified. You have technical policies and you have adaptive policies. Technical policies are relatively easy to implement. You can distribute money to the states. It's a policy. You can distribute it fairly evenly to the states because we've got a census that allows us to evaluate whether we've got poor people or rich people within a state and so on, so we could target money more to the poor. That's a technical policy. If we want to change what goes on in a classroom; if we want to create a new way of teaching reading or a new way of teaching math or whatever, that's an adaptive policy. That policy is going to play itself out very differently in the 96,000 schools around the country and, I guess, well over a million classrooms. To get there, it is going to be adaptive for a bunch of different reasons.

One is if it's done from the federal level, of course, it goes through all these layers. You've got the legislative layer with the compromises made. You have regulations written by the Department of Education with compromises made. You have the state departments of education all making their own judgments about what those regulations and that legislation says, and then they write guidance that becomes biblical, almost, to the districts. But the districts often change the bible through a rewrite of it and send it down to the school level, and so you get down lots of mixed messages. At the same time you have to adapt a single policy, a single line of policy to all the variation that exists throughout the country. There really is a one-size-fits-all mentality that goes on at both the federal and the state levels. That one size doesn't work.

The other point in here that I think is really important is that the ministries in many other countries where there is a "national curriculum" are generally run by educators. The decisions are generally made by educators, by people who have been in the schools. Often these are folks that have gone all the way up through the schools, been teachers and then principals, and then

kind of head of a district, or maybe an HMI in some of the English-speaking countries – Her Majesty’s Inspector. In the United States those policies aren’t typically made by educators. They are made by a mixture of politicians and lobby groups. It’s not the same thing. You can’t make the legislation as coherent, as together, because everything gets modified as you go through and you are making so many different compromises.

So the original intent here, of standards-based reform, was to create a coherent system with content standards being the driver; content standards as the goals of the system. Then you had alignment. You had alignment of policies, you have alignment of assessments, you have alignment of accountability around those standards. That’s the basic idea, but you would also have some sort of restructured local and state governments, and you would have the resources, flexibility and responsibility to meet the student standards. So that is the “opportunity to learn” standards. So these three sets of standards, now, would interact and work together to form a system where the accountability was fair. The standards were clear. They provided guidance where there was professional judgment made about modifying the instruction to meet the needs of students. Where you would have more personalization, it is called individualization in the United States. It’s called personalization across the world. The idea is that you create feedback loops that students need special attention. You need to continuously evaluate where students are and give them that personalized, individual attention as they go along. That’s the way you begin to get your equality into the system. Use a lot more tutors in schools where the kids need them a lot etc., etc. There are a variety of different ways to do this.

There also was a notion then, which was never put into any law, about ideas of professional accountability; not just performance accountability, but professional accountability and reciprocal accountability. These are different kinds of accountabilities but you need that whole cluster of accountabilities in order to make the system really work.

Ok, so the Clinton reforms come along in 1994. They were actually passed and signed in 1994. If you look at the literature on how does research enter into policy, one of the big ideas in that literature is that it is a big idea, that there is a big idea that gains credence over time. It gains credence because people hear about it, it makes sense, there’s a good storyline with it. The storyline of alignment is a good storyline. Everybody is always complaining about the discordance, the confusion of the system, so you create these storylines. You begin to create a package. Somebody looks at it, modifies it a little bit. Another package comes out, and after a while you begin to get a consensus around the set of ideas, and the consensus was fundamentally around the idea of standards along with alignment and with accountability.

Goals 2000 was brought in as a way of kind of priming the pump. It’s money out to the states to do things like create state standards-based reform. I know all that language was in there. States did a lot of different things with it, but it was kind of open money to them. They could use it for lots of different purposes, and that was a very specific strategy because people knew, people at the federal level knew that it was going to be very hard to get some states on board. It was very hard to get some states on board. In the first year I think only thirty states took the money. Then it became forty states, and so on. Mike Cohen flying around the country going to state legislatures and making arguments and going to governors and so on, literally going to

practically every state that there was trouble in. There were a lot of western states, a lot of southern states where there was a lot of trouble.

So you have the beginning of a structure that came along, and right after that IASA was passed. The idea in IASA was that you had to create a standards system for the poor but you also had to create the same...that if there was a state standards system the IASA had to be under that state standards-based system so that the kids who were poor were also served by the same state standards-based system. So it was a way of beginning to move the states toward having a common standards-based system.

I mentioned the implementation of the system. It's a really important part. I actually believe it took a bad rap because...Lynn actually wrote that by the year 2000, or actually in data collected in 2006, the vast majority of states had made substantial progress on the outlines of this particular reform. The implementation issue...to get it done in six years, I think, was really quite extraordinary given the situation, given the context. This also applies to, I believe, it applies to No Child Left Behind. It has only been five years, actually five or six years since No Child Left Behind came. We can't expect it to be fully implemented in that amount of time.

So NCLB comes in. A huge amount of concern by federal lawmakers, I think, frustration about the results of NCLB, of standards-based reform. The natural instinct of the lawmakers was to get tougher. Frustration leads to toughness. Frustration leads to striking out and trying to make something happen. Basically that is what they did. They took the basic structure that was in the IASA of the standards-based reform. They took that structure and they pasted on a whole bunch of testing, two to three times the amount of testing. They pasted on stringent accountability requirements. They pasted on ways of teaching reading and so on, and expected that all to be implemented suddenly and perfectly, and believed the theory behind it. That is, they believed that if you got much tougher in accountability those teachers were going to work a lot harder. Not a very good theory, in my view, that's worth arguing.

Ok, so what are the achievement consequences of it? We are going to look at two sets of achievement consequences very quickly. The first one is saying ok, does standards-based reform work? So prior to 1994, what was achievement growth like, and after 1994? So what we're looking at is changes in the rates of gain.

Nine year olds and reading. Between 1975 and 1994, as a nation we gained one point for nine-year-old reading, one point. That is a gain of .05 points per year. Between 1994 and 2004, over that ten-year period of time, there were eight points gained, about one point per year, .8 points or so. Take a look at blacks and Hispanics. Blacks gained .3 points per year between '75 and '94. They gained 1.5 points per year between 1994 and 2004. Ten points is a grade level, so between 1994 and 2004 African American children gained about a grade level and a half for nine-year-old reading. Hispanic children gained almost two grade levels. If you are looking for the thirteen-year-olds you see the same kinds of results. The gains are a lot more in reading after 1994 than before 1994. This is pretty important because it begins to give you an idea that, in fact, standards-based-reform worked, or at least there is a correlation there.

Now math is a little bit different. In math there were solid gains from 1994 on, but there were also solid gains before. So now let's go to between 1994 and 2002, and then after 2002. 2002 was when NCLB was passed. Again, in math you see the same rates of gain for both groups; serious, solid gains for both fourth graders and for eighth graders.

But in reading you see something different. So in reading what you see, if you look at whites here, now we are looking at nine-year-olds for fourth grade reading, they both gained about the same rate. Both gained about .6 grade levels, .6 points per year. But African American kids gained almost double during the period of time from 1994 to 2002, than they did from 2002 to 2007. Hispanic kids gained about fifty percent more. So their gains were greater. And if you look at eighth grade, what you find is even more powerful. You find gains of about a point for each of the groups; a point per year between 1997 and 2002. It's only a six year period of time compared to 2002 to 2007. It's either a negative gain or, at best, a gain of .2 points per year. So what you see is in math the same rates of gain and in reading very different rates of gain. Both of the rates of gains in reading in both grades, it favors before NCLB was passed.

That leads us to four questions. One set of questions...I think we've got to begin to debate a little bit about what the federal role can be, what the state role can be. We have to begin to think about what the differentiation in roles is. So how can the federal government support greater equality of results, support school improvement? How can it be supportive rather than prescriptive? This is the issue of balance in the governmental structures. Suppose that the federal government stopped requiring standards-based reform, including accountability. What would happen in the states?

Third is the world has changed dramatically since 1994. The word "technology" has not been used here at all in the last three hours, not once. There wasn't any web in 1994, in 1992. I gave a talk to a bunch of graduating seniors yesterday and we talked about what has happened in the eighteen years that these kids have grown up. There is all these extraordinary things that have happened and yet we are trying to design now, as a collective "we," we are trying to design something for 2008, 2009, 2010. We are using exactly the same kinds of thinking that we used in 1992 and the world really is very different. It is different in so many ways. You just think about the economy, you think about the communication, you think about the nature of technology, the delivery of technology.

Somebody mentioned that you can only change if you get down to the classroom level; the relationship between the teacher and the student. You can only change achievement if you are down there. I think that is very true. You can change. That is the place to go, but think about achievement more generally. People are learning all the time from technology. How do we begin to harness that in the schools, and outside of schools, around academic issues, to really change the nature of what we are doing?

Finally, we have learned a ton about learning in the last twenty years. We just know a lot more about how to teach, about how to instruct, about what the feedback loop should look like, about how to personalize all of this. We don't have a way of capturing that knowledge. One very powerful way that the federal government might be incredibly useful is to begin to put that knowledge up and available to people all the time in as unbiased a way as it could possibly make

it. So those are big questions. I have my own answers to these questions, but I would love to hear your answers. The next two or three years, I think, are going to be very, very important as we begin to think about the reauthorization of NCLB, or the reauthorization of Title I. I think about it as Title I.

My own view is that we shouldn't start with the existing law. We should start with a conception about what we want our children to know and be able to do, and what we want our schools to look like. And it's not just in math and reading. And it's not just in history, science and so on. It's as citizens that people are grappling with the problems of the world that are unbelievable at this point. Global warming is going to affect these kids, your children, in a way we can't even imagine now. Our kids have got to be able to grapple with those things. So we ought to start with some conceptions about where we want to go and then begin to figure out how we get there and what the different governmental responsibilities are. So I will leave you with that thought. Thank you.

Lynn Olson: Thanks Mike. Carmen, one of the traditional roles of the federal government, as Mike noted, has been to be aggressive in supporting the rights of underserved students, minority students, those who have not been well-served by state and federal governments. What has been the impact of state standards-based efforts and that of, first, as Mike mentioned, IASA and No Child Left Behind on limited English proficient/English language learners, in particular?

Carmen Perez-Hogan: I am going to talk about that impact as it relates to New York State specifically, and the work that I did in New York State when I was the coordinator of the Office of Bilingual Education. But before I get to that, a thought came to me, Mike, when you talk about the implementation problem, and I have to always relate things to my own Puerto Rican culture and language. In Spanish we have...in Puerto Rico we have a saying that says *friendo y comiendo*. That means you're doing things, planning and doing at the same time, and that exactly describes it in my mind, exactly what you were talking about, and that's really what has been going on in implementing Title III not only in New York, but across the nation and in NCLB. We had to be *friendo y comiendo* continuously.

The limited-English proficient students in New York represent the fastest growing segment of our population, with about 200,000 identified each year. They come from different countries, speak one of two-hundred different languages, enroll every single day at every single grade level, and they bring with them educational backgrounds that may be excellent, to youngsters who have very limited or no formal educational background.

Despite all of this diversity, this sub-group classification is based strictly on their lack of English proficiency. The impact of federalism on their education had its inception on the civil rights movement in the '60s with the recognition that the "sink or swim" generation of LEP students was not being educated. However, even before the *Lau v. Nichols* Supreme Court case, the New York State Board of Regents had issued its first Regents position paper on bilingual education. The paper described the needs of LEP students and the commitment to provide them with an adequate and appropriate education. Prior to NCLB, federal grants for programs for LEP students were going directly to the districts; and the SED, under my leadership to a large extent, provided technical assistance to help districts, and that was our role; and also to help them come

into compliance with the Lau remedies, and also to implement bilingual education programs under Title VII. And that was a lot of fun.

Then Title III of the NCLB – No Child Left Behind – really triggered some very dramatic and immediate changes, *friendo y comiendo*, on how we did we did business. Our role at SED changed overnight from being providers of this technical assistance to full responsibility for the implementation of the Title III NCLB program including the allocation of funds, approval of district plans, and managers of a brand new accountability system, *friendo y comiendo*.

A positive result of all of this was that the conversation on the LEP students expanded to many, many offices, in fact all of the offices at SED. This was a very good thing. It was no longer just out of our office that things were being done for LEP students. Title III also revolutionized the education of LEP students resulting in wonderful education opportunities, many challenges, and some disastrous student outcomes. It required that all LEP students be identified, that their progress in acquiring English proficiency be measured annually with a valid English proficiency test, and that they be included in the state's accountability system, and that they also meet annual measurable achievement objectives. Ironically, it is the only sub-group of students under NCLB that is currently required to be tested twice annually in English – once to measure English proficiency, and again to measure proficiency with another instrument, to measure performance in English Language Arts.

In New York...I am going to have to summarize a lot of this stuff because of time. But in New York we started working on education reform way before NCLB. The New York strategies for raising standards was approved by the Board of Regents back in 1996, which required higher standards in the core subjects for all students, including LEP students, assessment in the core subjects in certain grade levels, high stakes high school graduation requirements that required that all of our youngsters, including the LEP students, pass regents examinations in five different areas. Fortunately, those were made available in five different languages, including, among those tests, passing an English Regents test for our limited-English proficient students that created tremendous objections throughout the state. That is still in place even though strong recommendations were made for alternative procedures for late arriving limited-English proficient students to be able to meet that particular requirement. That requirement is still in place.

I am going to skip some of my notes. Maybe I can discuss it in the small group because I want to talk just a little bit about the archives and my experience with the archives in New York. I have been sensitized – and I wish that I had known what I know now that I am retired from SED – about the importance of the Archives and really saving documents. I must confess that I am bilingual but I am not archive proficient. I did, however – and I have good instincts – I did, however, for twenty years, file that work at SED; file, fold, and keep copies of almost every single little piece of paper that came to me and that I processed in any way, because I planned to write the book about the history of bilingual education and the education of LEP students for our students in our state. All of the rest of my paper will be in my book that I can't do today.

When I retired three years ago I took home copies of some of those files in lots and lots and lots of boxes, and instinctively I archived them in my garage. They were there for a while until they

were rescued by the SED archive detectives, John Suter and Jesse Brown. The lesson to be learned here is that in cleaning the office, my former office back at the department, for my replacement, what happened was that many of the originals that I had left behind were thrown out or misplaced or we can't find them anyway. I really think that we need strong guidelines out there, at least we do in New York State, on the importance of how to protect documents, which ones need to be protected, what we should do with them and about them. I wish I had known that when I was back in the department and that would have helped.

I have several recommendations. We need more for NCLB or whatever reauthorization is in place. We need teachers. We need teacher education programs. We need more funding if we are going to have all of this responsibility. The states need the money to be able to implement these programs and we can't afford to continue the policy of *fiendo y comiendo*. We need time to plan before we implement. Thank you very much.

Lynn Olson: Carmen gave us one particular state story. You've done research looking all around. And so we wanted you to step back and give us a broader picture, sort of, as federal law rolled out, what was the impact on states? How did they react? Also because we tend to get somewhat parochial in the U.S. about our own particular approach to standards-based reform, what might we learn by looking globally at how other nations have approached a similar set of issues or problems?

Marguerite Clarke: Thanks to Mike for giving a great presentation and to Lynn and Carmen. In my previous life as an academic I did a post-doc with Carl Kaestle, which involved looking up the topic of state responses to the No Child Left Behind Act. I will just briefly summarize what I did and what I found. And the findings are very much in line with what Mike presented in terms of the long term trends of states on NAEP, and how that has or has not been affected by No Child Left Behind.

For my study I wanted to look at all fifty states. I didn't want to look at just one. I wanted to see the whole federal, or U.S. landscape. I wanted to try to see what states' responses were to this major federal law and also to try to anticipate the effects of that law on student achievement in each state, and across every state.

So here is what I did. I started with states' policy starting points at the time No Child Left Behind was signed into law, and I gave each state a score based on the degree of congruence between its existing standards-based reform efforts and the requirements outlined in No Child Left Behind. So basically I was mapping every state's standard-based reforms effort onto the No Child Left Behind template, and if you got a higher score it meant that your pre-existing policies were more similar to the federal law. If you got a lower score it meant that it wasn't similar. Guess who got the highest score? Texas. It is a great example of states' impact on federal policy. I should have stopped there.

So I used those scores in two ways. The first way was I used it to group states into three groups: low, moderate, and high congruence; and they were used to represent implementation contexts. I used those groups moving forward to see if states that were very different from what the federal

law was requiring – were they going to respond differently to states that were more similar already to No Child Left Behind? So that was moving forward from 2002.

I also used states' individual congruence scores looking back in time. I looked at their performance on NAEP, on the state NAEP, in the six years leading into No Child Left Behind in a regression model, prediction model, to see where states that already had No Child Left Behind-like policies – were they able to make bigger gains on NAEP? Were they able to close the achievement gap? So it was like a pre-test of No Child Left Behind. I tried to see, is this going to be any better than the varied, multi-faceted standards-based reform model that currently existed in 2002?

Ok, cut to findings. My first finding was that regardless of state starting points, they pretty much all were moving towards implementing the No Child Left Behind requirements at the policy level, in the first few years; although this created significant difficulties in some states, particularly if they were rural states or if there were budget issues, but they were all still moving towards this No Child Left Behind model. So it was a good outcome for this more demanding federal approach, you could say. However, when I looked back in time at how states were performing on NAEP, I found that states that already had No Child Left Behind-like policies weren't doing any better than the other states. You didn't see bigger gains in them. You didn't see a closing on the achievement gap and I think we've seen since No Child Left Behind was passed into law that that has played out, particularly in terms of the achievement gap. The achievement gap is very resistant to these kinds of policies.

My conclusion. The conclusion I ended with in my chapter was that given my findings, that maybe it was better to allow more state variation in the absence of having a rigorous tested model that the federal government could require all states to apply. The conclusion that I personally came to that I didn't write in my chapter was I will never do a study like this again. It was so hard. It was so hard to find the data and even when you did find it, every state had a different definition, and thank God for Lynn Olson and Quality Counts. That saved me, to some extent. It was very hard to do any kind of cross-state work on this topic. I think moving forward it would be great to have some kind of common definitions, a central database would be wonderful, you could just go to.

Let me just move to lessons from other countries. Is that ok? I'll do it really fast. Mike already mentioned some experiences in other countries. I want to focus, in particular, on low and middle income countries, which is what I work with at the World Bank. I think what is or is not happening in those countries highlights some of the factors that may be underpinning the current state-federal relationship in the U.S. I think in a lot of lower and middle income countries you don't see the lower units, whether it is a federal system or not, the lower levels of the system do not really have an impact on federal or national level policy, which I would primarily relate to the capacity issue. I think capacity needs to be thought of not just in terms of, "Does a state or a local district have capacity to implement federal policy?" but, "Does it have capacity to influence it?" I think in many low and middle income countries they don't.

Another piece related to that is that the factors that do influence the kind of assessments and accountability models that these states have come external to the country. U.S. states don't just

impact federal policy in the U.S. They also impact what many developing countries are doing, particularly in the area of accountability models. I hear time and again ministries asking me, “What are states doing in terms of rating schools? Do you know of any good models from the U.S. for how states come up with ways to rate teacher effectiveness?” I think it is interesting that U.S. states are being looked at in that way outside the U.S.

Ok, I want to end with the concept we have mentioned often today and that was the concept of fragmented centralization. When I was reading Lorraine’s paper I was in Mongolia, which does not have a federal system, but I was struck by Lorraine’s elaboration of this notion of many different policies being implemented by all these different units, and it was very sophisticated and complex. I had just come back from collecting data for the grade five national assessment in Mongolia and it struck me that this kind of country was a great example of coherent or unitary centralization because when we were out collecting data in Outer Mongolia (I have been there) not only were the parents delighted to fill in the background questionnaire for the national assessment, but they also wanted to know when were we going to fix the roads and what were we going to do about the power supply. So to them we represented not only the Education Ministry, but also the roads, and maintenance, and electricity. So I think it again speaks to the point that if you don’t have capacity at lower levels, so much devolves to the central unit in a country. I think capacity is a key component of understanding how states impact and why they impact on US federal policy. Thank you.

Lynn Olson: We are, again, going to turn it over for table discussions and I believe we have until 12:20 for table discussions and then we will again try to gather up some of those questions that have arisen from your tables for up here, at which point those questions will only stand between you and lunch. So have a productive twenty minutes. This raised lots of questions for me, certainly, so I’ll dive in if no one else does, but have a good discussion. . .

Question and answer period following table discussions

Lynn Olson: . . .oving mics, can the people with roving mics sort of wave their mics around so people can see where you are? Ok, so we have one back there and I think we should have one up toward the front. Ok, and we are just going to go around quickly to a couple of tables and get some questions out for Mike and then the rest of our panelists, if they are sitting at a table, can also weigh in. So, first question, do we have one from back there? Great.

Question from audience: What can we do to follow through on Mike’s suggestion that we not start with No Child Left Behind, but go back to defining what we want students to learn? Should we focus on the LEA level rather than the SEA level?

Lynn Olson: Mike, quick response to that?

Mike Smith: I don’t have an answer. I think this is the kind of setting where we really need to have a, loosely called, a national conversation about this. I don’t think this can be led from Washington. I think it has to be led from other places like the [National] Governors’ Association, like the Chiefs, like a variety of the other organizations that actually reach out and touch the

states, maybe a Great City Schools, and then begin to bubble it up because Washington is not going to listen, I don't believe, until they hear it from people outside of Washington. Lots of you are from Washington. I am from Washington for fifteen years. I think Washington gets itself into a trap. It listens to itself, and reinforces itself, and then comes in with something that's a tiny little marginal change in whatever the law is, as a way of addressing law, and then goes home and goes to sleep happy that they have done something about this problem and they haven't done anything about the problem.

There have been a few times when Washington has moved in big ways. I see Bill Taylor shaking his head. One of them was obviously around the civil rights movement and the ESEA in 1965. ESEA did not come out of Washington. ESEA came out of a lot of research that had been carried out. It came out of people who came to Washington from other places, people like John Gardner and others. At that point he came out of Carnegie Corporation. He had been listening to a lot of people thinking hard about poor kids in the country. He had seen the research, and he and some others began to put it together. The same thing is true in the basic civil rights movement started by *Brown* – in some ways it actually started a lot earlier and worked up through. So this whole idea of a big idea moving into Washington; you are not getting big ideas that come out of Washington. So that would be my simple answer to it. It is not simple to do.

Lynn Olson: I saw we had a burning question at table six and also I know table three was waiting to get in.

Question from audience: We have a bunch of questions at this table that all kind of converge around this question of innovation. Where does good innovation come from? How do you get innovation without getting just a lot of bad stuff?

Lynn Olson: So how do you get good innovation and not bad innovation?

Mike Smith: I don't want to be the only one talking here. Just one thing. Lynn and I were talking about this, and this is something that the federal government really could do. It has to be done without an ideology around methodology, without an ideology around the right way to teach, without an ideology around the issue of ELL kids or disabled kids. It's got to be done in a transparent and independent way.

It is very simple for the federal government to do, to put \$300 million, \$500 million into serious work on R&D on the bridge between what we are learning about the brain, what we are learning about cogno-science, what we are learning about the use of technology and its application into education. This is a slam dunk. Any one of us sitting down with one other person could, in a day, put together an outline of an agenda that would be incredibly powerful and would change the nature of this country in five or ten years. But we don't do it. Instead, we bring our ideologies to the table, and then we argue about them and compromise them, and it ends up without very much happening. So how is that for an upper?

Lynn Olson: I think we have another question at table three, I saw, and then we'll wrap it up.

Mike Smith: Dave is going to do it right. We will ask Lynn to do it.

Lynn Olson: Ok, did you remember which one it was, Kathleen?

Kathleen Roe: Why do you think NAEP scores have flattened out under NCLB after rising between 1994 and 2002 under state accountability systems? Were the gains under state systems a result of new research in reading?

Mike Smith: I doubt there was new research in reading, certainly. You mean the earlier gains; you don't mean the later gains? I don't think so. I think there were two or three things happening. One was a little more alignment so the resources were a little bit better focused on achievement. I think there was a lot of talk about instruction, I mean, we were talking about what Tony Alvarado was doing in District Two [of New York City]. We were talking about what other groups were doing in other districts around the country. There began to be some talk about what is going on in North Carolina? What is going on in Texas? What is going on in Connecticut? Here is Connecticut exploring issues around professional development and really using that as the lever into their strategy. In the meantime, what is happening in Massachusetts? You have Massachusetts kind of moving along in a very interesting way and now showing up. If you looked at the mapping of Massachusetts eighth grade math scores they beat every country, except a few of the Asian countries. They beat every one in the developing world except a few of the Asian countries. This is extraordinary. So you've got these great victories that have gone on, and I don't think that there is any magic bullet in Massachusetts. I think it is a lot of hard work, a lot of alignment, a lot of thinking about improving instruction and so on. I think what happened was that you began to get this kind of activity going, up until about 2001, 2002.

Then you had a pin stuck into that balloon, or it was deflated because suddenly everybody was saying, "You're doing a lousy job. We've got to get in there and really kick some more ass, and be very tough about what we're doing." That frustrated a lot of people, and you saw it all over the country, and you talk to teachers and you see it, frustration. I mean there is a huge amount of psychological research on this and motivational research on this. If you set the bars so high that they cannot be jumped over, people don't try to jump over them. If you set the bar a little bit below that, at a level where it is manageable but challenging, that is the place you want to set it. You want to be really thoughtful about setting it, and it has to be transparent and it has to be, I think, it has to be a performance that means something to parents. It can't be a cutoff score. They don't know what that means. So I think we got hit over the head. The education establishment got hit over the head. People said that they weren't doing their job, they've got to talk faster. Al Shanker used to get asked on radio shows how do we raise the achievement scores? How do we do this and that? He would answer something about, "We'll get the teachers to talk faster," and everybody would laugh, right, but that's what we've been asking our teachers to do over the last six years. We want them to talk faster.

Lynn Olson: So I think it is actually time for lunch, unfortunately. This was a great session and I think it raised even more some of these sorts of tensions and paradoxes that we started off the morning with. The one that kept coming to mind to me was Lorraine saying policymakers gravitate towards high stakes tests because it is a lever that they know they can pull into the classroom. Mike is saying boy, but one of the most destructive things we've seen is this focus on high stakes tests, which then raises the question, well, if that is the lever what could the federal

government start to think about doing in this area that might be different moving forward? We are still going to be doing testing. We are still going to be doing accountability. Is there a different way to think about that? So lots of interesting paradoxes around that as well. Lunch is down the hall to the left, the same place we were for breakfast. We have forty-five minutes is what I have been told.