UNITED STATES OF AMERICA

+ + + + +

DEARTMENT OF EDUCATION

+ + + + +

THE USE OF SCIENTIFICALLY BASED

RESEARCH IN EDUCATION

+ + + + +

WORKING GROUP CONFERENCE

+ + + + +

WEDNESDAY

FEBRUARY 6, 2002

+ + + + +

The conference was held in the Barnard Auditorium at the United States Department of Education, 400 Maryland Avenue, S.W., Washington, D.C., at 9:00 a.m.

PRESENT:

Susan B. Neuman
Laurie Rich
Valerie Reyna
Lisa Towne
Michael Feuer
Stephen Raudenbush
Russell Gersten
Eunice Greer
Judy Thorne
Becki Herman
Linda Wilson

C-O-N-T-E-N-T-S

P-R-O-C-E-E-D-I-N-G-S

9:05 a.m.

SUSAN NEUMAN: Good morning. My name is Susan Neuman. I'm Assistant Secretary for Elementary and Secondary Education. It's just thrilling to have all of you here today.

One of our goals today -- we have a very practical goal actually. We're no longer debating whether scientifically based research and scientifically based evidence is important, we know it now is important and we know it is critical. As many of you know, we have counted one hundred and eleven times that the phrase "scientifically based research" is in our new law.

What our goal today is, is a very practical one. What we want to do is begin to explore the logic of scientifically based evidence or research and to really to begin to understand both its definition as well as its intent.

The second goal is something that is very particular to our office, the Office of Elementary and Secondary Education, and that is, how do we begin to put this into practice? How do we begin to suggest guidance?

What you are going to hear today is not only some wonderful papers on what is scientifically based evidence, what is it in its logic, it's characteristics, what it is and what it isn't. But, then, after a break, what we hope to do is really focus on what does this mean for safe and drug-free schools, reading, math, comprehensive school reform?

What we want to do eventually is move this debate throughout all of our programs so that we begin to really look at the scientific basis underlying what we say and what we do for schools in districts across the country.

What I want to do today is I want us to keep very much on pace. You'll see that there is opportunity to ask lots of questions. We ask you that the

questions you raise, please focus on the implications of this issue, not whether or not scientifically based evidence is a good thing or not.

I'm going to keep people very closely -- Valerie reminded me that I was already late. What we are going to do is we are going to keep people moving in a very fast pace and then give time for your questions. Then have a little break, move it on to implications and then, finally, have a panel where you really are able to address even more questions. We are delighted to have you all today.

What I'd like to now do is introduce Valerie Reyna. Valerie is the Deputy OERI, Office of Educational Research and Improvement. Her topic is what is scientifically based evidence, what is its logic?

VALERIE REYNA: Thank you very much. If you could go ahead and put my first slide up that would be great.

Welcome, it is a please to have the opportunity to talk to you and I gather that our well-organized organizer is going to keep the question and answer period to the end after all the speakers.

My usual style as a teacher is to have questions during the talk, so that's kind of constraining for me but I will try to contain myself.

MS. NEUMAN: You will be good!

MS. REYNA: Absolutely! But if there is something that is burning that's informational, if there's something that doesn't make sense at all, it wouldn't be a good idea not to communicate. So, please do raise you hand for that. At the end, of course, I will be delighted to entertain questions. In fact, a kind of give and take session is what I am really looking forward to, so that I can learn from you too.

Yes, that's who I am. We can go to the next slide.

I am going to talk briefly about: why scientific research, although I

don't think in the very short time that I have available that I could really give you a coherent argument that supports and defends the notion of scientific research, but I can touch on a few ideas very, very lightly.

One of them is: why scientific research? I think to think about that it's useful to think about what is the alternative to scientific research? If you didn't base practice on scientific research, what do you base it on?

Those alternatives include (this is not an exhaustive list, of course), it includes such things as tradition -- this is the way we've always done it, for example, superstition, there are -- you know, you throw the salt over your left shoulder and the reading scores go up! No, actually, there are things that are not based in fact that in fact become lore that if we really knew the scientific basis of it we would discover that those things in fact are just superstition. They are unfounded beliefs.

Then, there's anecdote. A fairly well-known obstetrician physician asked me once, "What's wrong with anecdotal evidence?" I think it is really a good question. Anecdote is a series of stories that you tell about things that have happened to you in your life. They can be very entertaining anecdotes.

The reason why we can't base practice on mere anecdote, however, and this is, of course, well known in medicine, is that individual cases may be exceptions. That may be the only case of that type.

In fact, anecdotes are often more entertaining when they are unique. But that is a weak basis to generalize to many, many people.

We know on the basis of experience that anecdotes have turned out to be false and misleading. Sometimes they are very representative, sometimes they're not. The problem is we don't know when.

Next slide. There's an analogy to medicine that I have obviously

drawn on already.

The first example, of course, is the classic one of when they used to bleed people. People would get sick. You know, I think it was when George Washington was bled that contributed to his death.

Why was it that good, well-intentioned physicians, because I think they probably were well-intentioned, I don't think they were trying to hurt the president, why is it that they didn't notice that it wasn't working? It wasn't just with this one patient, it was with many patients. Yet, somehow, personal experience was not sufficient to dissuade them from this practice.

Well, in fact, clinical trials are very recent in medicine. It was only in the 1940s that the randomized experiment where you know you had 2 groups, and you randomly assigned and all of that became routine and a standard, the gold standard in medicine. That is very recent in historical terms. Prior to that, we relied on those things I talked about in the first slide, like tradition and bleeding people.

One of the reasons why clinical trials are not sufficient has to do with the psychology of human thinking. I won't go into it in any depth, but I'm actually a cognitive psychologist and there's been research done about when you ask people to report about things they have directly observed and directly witnessed and the biases that can creep into that type of reporting. These are normal human biases that are generally adaptive, but they have predictable pitfalls. So, if you rely on your memory for past events, we know that that memory will be biased, and so on. Drawing simply on your personal experience alone is not a solid foundation for generalization.

Clinical trials in fact are the only way to really be sure about what works in medicine. The logic of it -- and the other speakers are going to go into far more depth than I really have the time to do, the logic of it is basically the following:

You have a group of people that you want to make a conclusion about. You want to say this intervention -- whatever it is, if it's a new reading technique, or whatever -- works for this group or not.

So, what you do is you take members of that population and you flip a coin essentially as to whether they are going to be in the group that actually gets the intervention or gets some kind of comparison, like what you would have done had you not done this new thing. Standard treatment, that's a common control.

The idea is that if you do this enough times and you get big enough groups, you've got two groups, the fact that you're flipping a coin ensures that these two groups, if you have enough people in them, are going to be comparable in every way except the intervention you're interested in.

Why is that? Because there was nothing that put one person in one group as opposed to the other. It was all by chance alone that you ended up in the reading intervention group as opposed to the control group. And, so, all the ways in which people do in fact differ, and people do differ, should be represented in both groups. They should be comparable in every way, except the one thing that you made different in their lives, therefore, we can isolate the effect of the outcome and trace it to that intervention uniquely.

This is the only design that allows you to do that, to make a causal inference. Everything else is subject to a whole bunch of other possible interpretations.

Now if you have too small a sample, obviously the logic doesn? t follow. Because you can have all the smart people in one group, the not so smart people in the other

if you only have a few. If you do this enough times, you get a big enough group, they will be representative. That has been proven mathematically by things like -- well,

we won't get into that!

The bottom line here is these same rules about what works and how to make inferences about what works, they are exactly the same for educational practice as they would be for medical practice. Same rules, exactly the same logic, whether you are talking about a treatment for cancer or whether you're talking about an intervention to help children learn. The same logic applies. In fact that's something I've said in talks for a period of time and the National Academy of Sciences report, which I know Mike and Lisa are going to talk about, in fact makes a similar claim. The rules of the game are the same.

I have the word "brain surgery" up there. The reason I have the word "brain surgery" up there is that I think, you know, when we talk about medicine and things like brain surgery and cancer, it is very, very important to get it right. We all recognize that and most of us buy into that. You know, that you've got to have randomized clinical trials because we want to be able to benefit for these treatments for cancer.

But when we teach students we really are engaging in a kind of brain surgery. We are effecting them one way or the other. Sometimes what we do helps, sometimes what we do, in fact, inadvertently, harms. We really don't know until we do a randomized clinical trial whether what we are doing is benefiting that student or not. We really don't know. It may be well intentioned, but that's not sufficient as we can see from the example from bleeding. So, it is brain surgery essentially and it deserves the same kind of respect for the nature of the consequences, in my opinion.

Next slide. So, I just told you that the randomized clinical trial, this randomized experiment where you can assign people to two groups and chance alone determines which one they end up in so that they are comparable in every way

except for that key thing you want to look at in terms of cause and effect, I said that is the best form of evidence, and it is. It is the best form of evidence.

However, do we have a lot of that type of evidence in this field that you can draw on? Now, we've exhorted you through legislation and a number of other things, you must use this, but is there a lot of gold standard level evidence out there about all the things we do on a daily basis in the classroom?

No, there isn't. There is some. There's some evidence out there. A lot of the evidence, however, is lower on the hierarchy of the strength of evidence. I am going to just touch on this briefly. Again, the other speakers are going to talk about it in more detail. When did I start?

MS. NEUMAN: Like ten of.

MS. REYNA: Okay. So, there is a lower level of evidence that we can describe as quasi experimental or large data bases that essentially have lots of characteristics of students in them that you can correlate with one another and you can correlate with outcomes.

The idea here is that nobody has been randomly assigned. In the real world randomness is a very rare thing. It's a very artificial thing. In the real world there's lots -- everything's correlated with everything else.

Think about the example of socio-economic status. Correlated with everything, you know, your neighborhood, your number of books in the home, all of these things are associated in real life.

But when you look at the pattern of associations, you can go in through statistical magic, that's basically it, and you can artificially create a sort of comparison or control by sort of equating people on things. If you look at enough different combinations of people and enough different characteristics you can statistically attempt to control, to capture basically the logic of that gold standard, the

randomized experimental trial. That's always the logic, that's always the goal.

But here you attempt to do that by statistics. It's not as good. It's a lower level on a hierarchy of evidence, because there could always be something you are not controlling for that in fact is causing your outcome. That's always possible.

However, it is second best. It is not nothing. So, for example, you at least know that something is maybe probably true, that there's a large number of what's called in public health epidemiological studies, and there would be an analogy in education to those large studies, lots of attributes, the obvious things controlled for. You know, you could at least say, well, it's probably true. That's certainly better than we have no idea, much better than no evidence, well, what do you think? It's not the top level of evidence, but at least it is evidence.

Another thing that is a good source of extrapolation to practice is evidence based theory, and the evidence based theory is the crucial part. Theories whose predictions have been confirmed and disconfirmed -- you know, there's been an opportunity to disconfirm them as well, they've been tested -- that are explanatory, that go into the mechanisms of how people learn, how they learn, what's the process going on.

If you know something about how people learn and how an intervention was effected, than you have some clue as to whether you can generalize it to your classroom, because you know the mechanism. You know what's relevant and what's irrelevant to the causal course of that intervention.

Is the shoe size of the student relevant? Probably not. Why is that? Because we have an inclusive theory of how learning happens and it doesn't have to include peoples shoe size. Right? So, if we have a tested theory, we can sometimes extrapolate beyond just the limited group that was originally studied. You

know, sort of the boundary conditions for when an intervention is likely to be effective.

Are there pitfalls of theory based extrapolation? Yes, because sometimes it can turn out to be that it doesn't follow for that group for other reasons that weren't study. So, there are always pitfalls.

A lot of people worry about the fact that science, in some peoples view, is a soulless, heartless enterprise. What about the student as a person? What abut the interpersonal relationship between professionals, teachers, principals, so on and so forth and the student? Doesn't science really take the heart out of things?

I would argue: definitely not. When you give students the opportunity to learn and be successful that supports them as people to.

Moreover, there is really no dichotomy between science and values, for example, or science and emotion. That is a false dichotomy. When we think about values, I think it is important to recognize that evidence does not determine our decision solely. It is not just the facts. It's the facts plus values. But without the facts, we might make the wrong decision, even based on our values. Because we don't know what's true and what's not true.

The facts, the evidence is necessary to make decisions that effect students' lives, but it's not sufficient. But it is necessary. That is what we're promulgating, that, at least, it be part of the discussion so that we can base practice on it. So, we're talking about science with a human face, and that's a person --.

This whole enterprise of translating scientific research into practice is very complicated. There is even research on how to do that. It's called translational research. In medicine, for example, there's a lot of that.

That last bullet there is really an invitation for your help. I am at OERI, the Office of Educational Research and Improvement. We are thinking very

hard about how to do this, how to most effectively be useful to you and to support you in what you are doing.

So, I would be very, very interested in suggestions that you might have. I am going to stay for the whole day and practical suggestions about education and

training, that sort of thing, would be enormously helpful for us. I think this symposium we have here today is a wonderful first step in that. But, it's the kind of step we need to take and we need to take a lot more.

Next slide. What is evidence based education? I am going to go through the next slides much more rapidly. I'm just going to sort of allude to points, and then if people want to talk to me more in depth, I'd be happy to do that. This is going to be pretty fast.

We can't get the slides up over there? Can you see and can you hear?

So, what is evidence based education? The best available empirical evidence in making decisions about how to deliver instruction. But, again, we don't have even the second level evidence about all the practices that currently occur in the classroom. Nor do we have even second or third tier level evidence about things that have to be accomplished in the classroom.

So what is a professional to do? That's when human judgment comes into play, to fill those gaps in evidence. That is inevitable. You have to apply your judgment. There are whole books written and research done just on the nature of human judgment. As you make decisions, you might want to dip into that literature. It's actually quite helpful. Leaders of industry and business often get consultants to advise them about the nature of decision making and decision analysis.

In a nutshell, what I would say is that there is a lot of wisdom in

human judgment. That has been empirically demonstrated. There is also systematic bias in human judgment. That's also been empirically demonstrated. It's an inevitable thing that has to be an ingredient today and probably for many, many centuries more.

We are just not going to know everything right now. That is the nature of science, and we are going to discover new things that make the old knowledge obsolete.

But, at least, in science it is cumulative progress. It builds on the knowledge of the past, if it's truly science. It doesn? t throw away things people have learned that in fact have been effective. That is not the nature of science. Science is by its essence cumulative.

What is empirical evidence? Well, the most important aspect of what's up on that slide, is that it's objective evidence. It's the kind of evidence that if two people watched something, they'd say yes, that's what happened there. The interpretation of that evidence has to do with what I alluded to earlier having to do with causal theory. That's a whole other level, but at least what happened at a surface level is agreed on. Then you make hypotheses about why it happened and you test those and you can be wrong in science. That's the nature of empirical evidence.

Scientific research really is evaluated primarily on two big dimensions. One of them is the quality, and that is primarily in terms of scientific merit, and that has to do with the method. When I was talking before about randomized experimental trials, and large correlational studies, that's methods, methods of analyses. That has a lot to do with the quality of the evidence. So, if it's high on the hierarchy, if it's the gold standard, it's top quality. If it's one notch down, it's second level quality and so on until you get to things that are really at the level of

anecdote which are maybe slightly suggestive, but they're not the highest quality of scientific evidence.

Relevance and significance, obviously, is the other criterion. Scientific merit and good methods alone don't make the best scientific research. It has to be relevant to your practice and it has to be significant. The more significant it is, the more people are effected by something, the more severe the issue is that's being effected, obviously the more important the research.

So, if you look at the National Science Foundation, for example, and you look at the way they evaluate grants that they receive in the sciences, it turns out to be exactly those two criteria: scientific merit, relevance and significance.

Next slide. So, here's a little bit more detail on what I talked about before about levels of evidence. What are the levels of evidence? Again, for those people who can't see, we'll make this available in some form or other.

Again, the other speakers will be talking much more in detail about his. But, we have our randomized trial at the top, then our quasi experiment, then our simple correlational study, and so on down the case studies.

Go ahead. This is the logic once again in more detail about why randomized control studies are the gold standard, why they're the highest level of evidence, why it's what you should rely on with the greatest weight by far.

Again, there's self selection bias operating in the real world. What that means is people are assigning themselves to groups in the real world and it's not random. People of a certain type tend to belong to certain groups to do certain things.

People who smoke tend to drink more coffee. So, is it the coffee or is it the smoking? Well, you have to control for the drinking of coffee. It's that sort of logic.

Next slide. Why is randomization critical? Because it equates on this ways in which people are -- differ that are correlated with one another. That's why it's so powerful.

Again, this is just more detail for a longer talk.

That's just an example. You can go ahead and skip that.

Now, when you think about relevance, this is a very difficult thing. Scientific merit you should use the hierarchy of evidence as your guide and that's fairly straight forward.

Relevance, on the other hand, is a much more sticky issue and much more difficult. But, one of the key things you can look for is does the study involve a similar intervention outcome to those of interest. You'd be amazed at how many times people say there's evidence for something, then you go look it up and some very obvious things are wrong like they studied something else.

They say one thing and it's really something else. So, they say, okay, the effect of the graphing calculator on the ability to, you know, do certain kinds of mathematical computations without the calculator, you know, there's some arguments about transfer. And they didn't look at graphing calculators, they looked at non-graphing calculators. This is common sense.

So, you'd be amazed at how many things you can screen out by asking some simple, common sense questions about relevance. You'll screen out a lot of the junk by doing that.

One of the things you can do is you can search the literature, obviously. Some of that requires, however, you know, folks that have advanced training. And how to do that and how to bridge that is something we should talk about.

You can screen. Obviously, you should screen on the two

dimensions we talked about, quality and relevance. Those should be your touchstones. You can search for evidence that has been interpreted. For example, I give an example of narrative reviews and meta-analyses.

However, when people summarize the literature and they say they are summarizing the research in a field, the quality of those summaries varies a lot. Some of them are essentially an opinion piece. This is what I think. People's opinions are interesting, but it is not something you want to necessarily base the lives of millions of children on with great confidence.

Some reviews are much more formal and meta-analytic and scientific and another person looking at the same literature would make a similar conclusion, those are the ones you want. So, meta-analysis is totally superior to a narrative one.

Go on to the last slide. This is the part where we talk about what we are trying to accomplish that we hope will support you.

These are our goals and they are in our strategic plan and we really mean them. We're trying very hard to achieve these goals.

We want to provide information and tools. The goals we are ultimately looking for here though is that, as it is in medicine today, that at some point and I think this point is inevitable in the future, at some point the use of scientific research as a basis for educational practice will become routine. It will become customary and people won't be able to imagine a time when that wasn't done as a matter of course. Thank you.

MS.NEUMAN: I think she makes that more clear than anything I've heard for a long time.

I'd like now to introduce Michael Feuer and Lisa Towne. They have just completed a wonderful project on scientifically based evidence. I am

wondering if you have that report with you?

MS. LISA TOWNE: I didn't anticipate to have to provide this many copies.

MS. NEUMAN: Lisa is a Senior Program Officer at the Center for Education at the National Research Council. Michael is the director of the Center for Education at the National Research Council of the National Academy of Sciences.

We are delighted to have them work with us in talking about the logic and the basic principles of scientific based research, as well as help us focus later on on the implications of this research for practice.

MR. MICHAEL FEUER: Thank you very much for this invitation, Susan, and thank you to all of you for coming out to listen to lectures about science on this Wednesday morning.

We're here to tell you a little bit about a report that was released at the end of November in this handsomely bound pre-publication form. It's called "Scientific Research in Education." I want to spend a few minutes telling you some of the highlights of both why we were asked to do this and what you would find if and when you opened the book and read it which I hope you do.

First of all, the National Research Council of the National Academy of Sciences is, as I'm sure you know, an independent organization. We are not part of the government, although we work closely with the government and on behalf of the American people. This is an idea that goes back actually to the 19th century when President Lincoln looked around and discovered that there were some serious problems that perhaps science and technology could help him with. I'll just tell you one quick story which my poor staff hears this so much that they tend to nod off when I get into this, but if they'll indulge me.

One of the very first problems that this new Academy was confronted with had to do with a problem in the Civil War which was the ironclad ship. This, as you recall from history class, was an invention that actually ultimately helped the north win the war.

There was a problem with the ironclad ship, however, and that is that they couldn't get the compasses to work because of the magnetic fields. Now, if you are ever interested in a sort of classic case of the collision between science and public policy just think about a ship that you can't get to -- you know, knowing the difference between north and south with the Civil War at hand is not a trivial matter.

This was one of the first problems that the Academy was asked to solve and, indeed, a small committee of physicists and engineers was brought together and they actually solved the problem, and I am actually happy to tell you that the report is nearly through review.

(Laughter.)

Now, with respect to education and education research, this is not the first time the Academies have been asked to weigh in on this. There were reports going back even to the late 1950s, and then later through the '70s and '80s and '90s. And that, in itself, I would submit, is an interesting little bit of evidence (perhaps anecdotal, but maybe not) of the very perception that education research has, at least in part, an important scientific component. Because, after all, we are not the National Academy of Poetry, we are the National Academy of Sciences, and when we were asked to take on a question of the scientific quality of education research, I don't think that was coincidental. I think that is part of a very important perception in the land about the nature of education research.

Indeed, when we were about to launch this study most recently, I began speaking with some of the distinguished scholars around the country. And,

when I mentioned that we were about to do a project on the scientific quality of education research, I have to tell you that one of these very distinguished scholars said, "Well, that's great. Finally, we'll have a short report from the academy."

That's another important perception that we had to deal with and that is that the general perception of a low level of quality in education research writ large.

We don't have any evidence and we didn't try to get evidence to support or refute the claim of the overall quality of education research being poor. But we did take as a datum that the perception that it is poor is important and that it is, therefore, worthy of the attention of some very distinguished scientists and educators to think about this whole question.

One more bit of context. I don't think it is coincidental that requests for study of the scientific nature of education research should come at a time when we probably have more information, more data and a more relentless flow of ideas about how to fix the schools than perhaps at any time in history. Again, I haven't done the empirical research on this, but I would bet that education policy gets more headline attention than almost any other item on the domestic agenda. To some extent, I think the Administration, and Congress have conveyed an incredibly powerful message in the passage of No Child Left Behind, in particular just after this horrible season of terrorism that we have just come through. It is again an indication of the overwhelming importance of education and education policy in the agenda.

But, that said, there are lots and lots of folks who have gone to school and who therefore have very firm opinions about how to fix the schools. What we get is a cacophony of ideas, solutions, reform initiatives, standards -- I mean, we're responsible for some of the standards documents. And, I sympathize with people in the real world such as yourselves and with teachers and educators all

around who have to sift through this morass and make something significant and effective. That's where the appeal of science becomes very strong. It is after all an enterprise that attempts to distill from the cacophony of ideas and anecdotes and impressions, the nuggets of really enduring value, and that kind of knowledge upon which you would want to base important decisions about kids, about schools and about, ultimately, ourselves.

Having said all that, let me just offer a little bit of a foundation here for what Lisa is going to tell you more specifically and that is some of what's actually in this report.

As I said, we are an independent organization. We were asked to take on a set of questions having to do, really, with first principles: What is science? That in itself took a few weeks to sort through. What are the principles of science and how do they apply to the science of education? These are very tough questions. What you will hear about is some of the key findings of an interdisciplinary group of scholars, not all educators: cell biologists, a chemist, education scientists, statisticians. This is the way we do our work. We bring these types of people together. And, after all, the National Academy of Sciences obviously exists in some measure to promote the values and the ethos of science and it's utility in public policy decisions.

So, much of what Valerie has said resonates with the underlying purposes and -- are we trying to follow along with the slide show? Because nothing I've said so far is on any of these slides. We have a unit at the Academy that specializes in improvisational theater.

(Laughter.)

Let me make this one little attempt at a slightly more cautious definition, or a more cautious statement about the nature of scientific reasoning in

education research.

On the one hand, I think what you would see in the report and what you'll hear about is a great deal of enthusiasm and encouragement for the notion of bringing scientific reasoning, the culture of science, to bear on the important decisions we make about kids and schools.

After all, science is intendedly rational, it is disciplined, it is honest, it is open, we aspire to a kind of dispassionate, politically neutral distillation of evidence to make decisions. That's why we are enthusiastic about the underlying proposition here that has been articulated in the law and that most of you now are going to have to turn into the real practical day to day.

At the same time, I want to tell you that what scientists themselves often acknowledge is that there is a dimension of human judgment that can be missed with an overzealous focus on the rigors of scientific method.

It was, in fact, a psychologist who won the Nobel prize, Herbert Simon (unfortunately he passed away about a year or so ago) whose contributions to this I think are quite significant because of his work on what human rational decision making is really all about.

The story that he liked to tell was about the traveling salesman who had the following problem: to visit 15 cities and to work with customers in 15 different cities and wanted to minimize the costs of visiting those customers, fuel costs, time and so forth. What's the rational way to approach that problem?

Well, one rational way to do it is to figure out the different routes you could take and then calculate how much it would cost because of the mileage and the fuel consumption.

Is that, however, really rational? And, the answer is not necessarily. And that's because by the time you lay out all of the different routes and

you mathematicians out there will figure this out pretty quickly that 15 factorial routes is a pretty large number. And, so by the time you have gotten to the end of the list, 20 years have passed. Your competitor who is using a less rigorous, less optimal approach has gone to Cleveland and then figured out that the next stop ought to be Buffalo because that's closer than Houston. And, you're back there on the back of your envelope doing the science.

The question becomes what really constitutes rational decision-making? And, the answer is: it depends on context, it depends on technology, it depends on the time you have, and, frankly, as Valerie has I think so eloquently reminded us all, a lot of the decisions that have to be made are going to be made with less than perfect evidence.

And, therefore, you have a double challenge. One of your challenges is to encourage the field of research to provide you with better and better useful evidence. And, don't think for a minute that we researchers have figured all this out and the only problem is you people in the real world aren't using it. We know that's not true. The research community has a lot to do to shape up in order to provide you with useful evidence.

At the same time, the challenge is to continue to make reasonably good decisions based on the evidence that you have.

I don't want to take time away from Lisa because the real messages of this report are what I think are going to count at the end of the day.

So, I thank you for letting me give you a little sermon about rational decision-making. And, now I will try to sit down rationally and let you hear the rest of this.

(Applause.)

MS. TOWNE: Hi, everybody. It is a pleasure to be here. I just

want to, with time considerations, just sort of pick up where Michael left off and like he said just give you a brief sort of tour through what's in a somewhat longer volume.

As Susan suggested, I am happy to make copies available to people. I wasn't able to bring them here today but I will work with her to make sure that we can do that and that you will have the pleasure and the privilege of reading every page.

In the meantime, what I am going to do is just talk through, give you a brief idea of what's in here with respect to the question we were asked to talk briefly about today which is what are the basic principles of science. As you might expect and as I'm very grateful to report, they do reflect in many ways what Valerie has already said.

MS. REYNA: Good.

MS. TOWNE: Yes, this is good. So, Steve, get ready!

If we could go to a slide that says, "Principle One," that would be great.

What I'm going to do, just to give you a sense of what I'm going to talk about today is talk briefly about the principles of science that actually are common across all disciplines and fields. This is, again what Valerie said, that at a fundamental level, medicine (that was the example that she used), ecology, economics, all of the applied fields like medicine and agriculture, that there is a lot that is actually shared between them.

The principles of science that I'll talk about today is what the committee who wrote this report believes are those common elements.

Then, I'll spend a few minutes at the end talking about what is it about education that makes the application of these principles look a little different. Because you might be sitting there thinking, "Wow, looking at something that a

physicist does sure does seem a heck of a lot different than what an education researcher does."

So, I'll talk a little bit about how these principles play out in studying education and why it is that they tend to look very different.

So, the first principle here relates to posing questions. It seems very straight forward, perhaps self evident, but actually the process of posing a new and different question is often times itself what is the basis of a scientific breakthrough, someone thinking about a problem in a new way and asking a new question.

There's a couple of words here that I'll just touch on briefly that give a little bit more detail about what this means.

"Significance," this again goes back to what Valerie was saying with respect to education. The significance of a question can be judged in terms of its relevance to the core problems of teaching and learning and schooling.

In a more traditional scientific sense, the significance of the question can also derive from what has come before it. In other words, does this question help to advance the field and consensus, and the cumulative nature of science which is a theme that Valerie touched on and that this report also tries to stress very strongly.

The second one, I'll just touch on briefly, is "empirical." That simply in very straightforward terms means can be observed. The only reason this is relevant here is because there are some questions that are relevant to what teachers do every day that can't be answered by science. Should students be asked to say the Pledge of Allegiance every day, for example, has to do with our values as a society and whether we think that is appropriate and good. It is not something that can be subjected to scientific study.

I will go on to the next slide, and talk about the principle that has to do with theory, and again Valerie alluded to this as well. The importance of theory is really very important in education research and the other sciences as well. In fact, much of science is fundamentally concerned with the development and testing of theories that helps you explain some aspect of the world.

In hard sciences, so-called hard sciences, we know of theories like evolution. Grand theories like that don't typically pop up in education but certainly they are relevant and they certainly are kind of an organizing conception for scientific work. Valerie mentioned a theory of how children learn, that's a great example. A theory of how educational resources translate into outcomes in schools is another example.

So, theory is really kind of an organizing idea for scientific investigation. The important point here is that data in an of themselves aren't really relevant to a scientific investigation unless they are related to some sort of conceptual idea that you have going in like about how children learn or about how educational resources translate into, hopefully, better outcomes for schools and for students.

Even in program evaluation, which is a lot of what has to do with the implementation of this law, what works, there is some implicit theory about how the program is supposed to actually translate into better outcomes for kids. Should that point to the basis of a program evaluation? That's what Carol Weiss calls "a program theory." So, sometimes it's explicit and sometimes it's implicit, but it's always there.

I will go onto the next principle on the next slide. This has to do with methodology, which Valerie has already, thankfully, covered very well for me.

I will just make three main points about the role of methodology in scientific research.

First of all, that there are a range of legitimate methods in the field. Education is studied from a lot of different disciplinary lenses: economists study this, developmental and popular psychologists, sociologists and anthropologists, they're sort of studying a different part of the animal and they all bring their tools of the trade to bear on that. So, by definition, there are a range of legitimate methods that are within this domain.

A related point is that when you're looking at questions in education research, that multiple methods used together tends to strengthen the inferences or the conclusions that one can draw when studying these things scientifically.

The last point that I will make about methodology and this gets to Valerie's hierarchy of evidence, is that although there is a range of valid and legitimate methods that can be used in studying education, some methods are better than others for particular purposes. Valerie, actually, kind of very nicely laid out kind of a hierarchy of evidence within the class of questions that are causal.

There are other kinds of research in education. There's descriptive research. There's research that looks at mechanism. And, within those classes of questions, there's also different kinds of methods that can be used. So that the method itself, taken out of the context of a particular study, can't really judge to be good, bad or indifferent. A method is only as good as it addresses a particular question that is being addressed.

I'll go on to the next slide. I have three minutes and I have several more principles.

Principle four is: a coherent chain or reasoning. This is sort of the logic behind science which, again, Valerie, has talked about and handled quite well.

So, I'll go on to the next slide which is principle five, and this has

to do with replication and generalization.

"Replicating" is a very core notion in science. It has to do with the fact that since in any particular study you're only relying on a limited set of observations, to what extent does what you're looking at here and now generalize to other times, places and contexts. In education, as you know, this is a critical question. Teachers and researchers alike have been knowing for years that something that works in a particular classroom may not work in the classroom next door and may not work in the same classroom a year later. So attention to sort of what's going on in the classroom at that time can help you understand the conditions under which things tend to work and therefore how to think about how findings can generalize from one time to another.

I'll go on to the last principle here, which has to do with the transparency of the scientific enterprise. Valerie alluded to this as well. This just has to do with the role of the scientific community actually working together to try and make sense of all of the findings and all of the conclusions that come from individual studies. Educators often bemoan what there perceive as bickering among the research community and we'll grant you that there is some bickering. But there is actually something important to say about that and that is that researchers are actually trained and employed and paid money to be skeptical observers and to ask critical questions. That's their job. So, this critical kind of work, critiquing other peoples findings and trying to make sense of them is actually an indication of the health of the scientific enterprise, not its failure.

So, those are the basic principles of what actually binds scientific inquiry together across domains and disciplines.

I am going to just touch briefly on a couple of things in education that help understand how these principles are actually translated in the study of

education. How much time do I have for that? One minute? I am obviously going to just whiz through these.

One issue has to do -- at one level there is a difference between the so-called hard and soft sciences. And, that has to do with differences that emanate from studying inanimate objects and studying people, which are complex and do crazy things that we often can't understand or predict very well.

So, there are some things that are different. Broadly, research or control is one of them. Think of it this way, a petri dish of heart cells is a heck of a lot better behaved than a classroom of third graders. Anyone whose tried to study education research and has done cell biology, as one of my committee members did, can attest to this.

There's other things that are different. I'll just touch on this last one on the slide which has to do with certainty. Valerie said, and the committee completely agrees, that science is by definition an uncertain enterprise. The key is understanding the degree of uncertainty that is associated with what we know. In general terms, in the physical sciences we because of this ability to control the environment tend to have more certainty associated with them than sciences that have to do with people, like education research.

Moving on to the next couple of slides, there's a couple of things in education, specifically, that actually explain and help understand the nature of education research. Values and politics, Valerie talked about this as well, the role of schooling in our democracy is one that is appropriately and historically grounded in our values as a people. What we decide to do with respect to schools is inevitably and appropriately going to be grounded in those values. Scientific research is one part of that decision process and it should be, but interacts in a very significant way with our values.

Human volition, I've alluded to this already. This has to do with the fact that people don't always have the same agenda as a researcher might and they might move around and mess up samples and do things like that. So, there's some messiness that researchers have to deal with.

Variability of education programs, I don't have to tell all of you about the differences in the implementation of programs that happens in different districts and schools.

And, the organization of education, the fact that we have sort of this nested hierarchy matters in education research because understanding what's going on in a school, you have to have some understanding of what's going on in the districts, in the state and even at the federal level to really have a good sense of what's happening at school.

Just go on to the last slide, there's a couple of remaining points that I'll just touch on and then wrap up, about what characterizes education research as a profession that tends to help understand its nature as a whole?

One is something I've alluded to already and that is the fact that education is not a traditional scientific discipline. It is an applied field, like agriculture, like medicine. So there are a lot of disciplines that legitimately bear on our understanding what is going on in education and that is a key piece to understanding it.

Ethical considerations. Most sciences, but not all have really to be concerned with the ethical implications of what they're doing. Studying kids who are a vulnerable population sometimes entails things that you have to do with methodology and plan for research in order to make sure that they're protected. Most of the time education research doesn't pose any risk and is exempt from the federal regulations that govern them, but, none the less, it is something that factors

into the research process and shapes it in a significant way.

Finally, I'll end with this notion of relationships. Researchers can't do their job without the cooperation of schools and students and all the different actors who are in the education system. At the very least, they need the cooperation for them to go in and collect data, to test them occasionally and increasingly we're seeing full blown partnerships being developed where researchers and educators who are on the ground doing education day to day so to speak, actually work collaboratively in a way that tries to both improve practice through research, but also inform and improve the research process by better understanding of what's going on in practice.

With that, I will conclude.

(Applause.)

MS. NEUMAN: There's nothing worse than feeling rushed. I hate to do that, but unfortunately we do have a lot to cover.

It's a special pleasure today to introduce Steve Raudenbush. He's a colleague of mine at the University of Michigan, and he's one of those methodologists that actually talk in human language.

He's a wonderful translator of research evidence and what we should begin to look for as we become critical consumers of research.

Steve?

MR. STEPHEN RAUDENBUSH: Thanks, Susan. Susan made me promise not to show any equations! I will have no slides, so if anything's going on up there, pay no attention to it.

(Laughter.)

In May of 1999, I had the good fortune to attend a meeting at the American Academy of Sciences, not to be confused with the National Academy of

Science. The topic of the meeting was how to improve the scientific quality of educational research.

The two main organizers were two venerable characters named Howard Hyatt and Fred Mosteller. For Mosteller and Hyatt it was a kind of a déjà vu because they had been among the most influential people half a century earlier in advocating effectively that medicine should be based more on scientific research. They felt that the time was appropriate to make the same argument now in education.

At the time they made the arguments with respect to medicine, they were met with considerable skepticism. There was a famous (at that time, at least) well publicized debate between Hyatt and a heart surgeon. Hyatt was arguing that we should do experiments to see whether new surgical procedures are really effective as compared to let's say medication. The heart surgeon asked him in a very poignant moment, "Sir, have you ever held the beating heart of a human being in your hand?" The surgeon argued that the cold logic of science did not replace the clinical judgment of the seasoned practitioner.

Hyatt and Mosteller, of course, in response argued that in a lot of cases the medical profession really doesn't know what the best thing is to do and that in that situation it is unethical not to find out, and in fact if we can find out what works best than over the many years many millions of people perhaps will benefit and that would reveal the true ethical character of basing decisions more on science.

Over the last forty to fifty years, their argument, that of Mosteller and Hyatt, has in many ways I'd say largely won out, that we now in fact accept and admire the commitment of medical professionals to base, not all certainly, but some of their key decisions on research from clinical trials.

One of the questions that comes up that's interesting is what caused the sea change in medicine and is it likely that anything like that might happen

in education. That's way too big of a question for me to try to answer, but there is an interesting vignette, I guess, a part of the story that has to do with the Salk vaccine for polio.

In the early studies in the '40s and early '50s on the Salk vaccine, the studies seemed to show basically that the vaccine wasn't effective. People who had the vaccine were almost as likely or it may have been in fact equally likely to get polio as those who did not. By the way, at that time the vaccine had not been perfected. It was certainly far from perfect.

But subsequent research showed that higher income families were more likely to get the vaccine and higher income families in this case were more like to in fact get polio. It transmitted in places like swimming pools, places where high SES people, social class people actually had a higher risk.

Subsequently in 1954 was a very important, huge, national, randomized clinical trial on the vaccine. This was a double blind trial in which physicians didn't know what vaccine, what treatment they were giving to people, whether it was actually the vaccine or just the placebo of sugar water. And, the people who were getting it didn't know what they were getting. Having grown up in that era, you have to realize when you got sick in those days and the doctor came to your house, remember when the doctor used to come to your house? (Laughter.) Your parents would stand by in mortal fear as the doctor exercised your legs and did various things to see whether it was polio.

So, here people were doing this double blind randomized clinical trial and the people didn't know what they were getting and the doctors didn't know what they were giving. It's quite remarkable that this happened.

But the results showed definitively that the vaccine was far more effective than not having the vaccine which led to further perfection, further clinical

trials and ultimately the wiping out of polio as a disease.

Now, we may not expect quite such dramatic success in saving lives in education, although the relationship between education and health is actually a very durable and interesting one, so maybe not being educated can cause a loss of lives.

But there are striking parallels in education. The first evaluation of the Head Start program showed roughly equal cognitive skills at the end of the study if you compare the Head Start and the non-Head Start kids. But subsequent research showed that the Head Start kids had higher levels of poverty than the non-Head Start kids. Some then argued that the results actually showed that the Head Start program must be effective because the kids were doing better than you would have expected them to do given their social background.

So, here's a result of two groups basically being the same and one group of people saying this shows Head Start is no good and the other group of people saying this shows Head Start is really good. The same evidence, but the evidence is so weak that it can't really decide the question. Unfortunately, there was no follow-up experiment to give us a better answer.

This leads to a crucial point that Valerie made. In both there are striking parallels, as I said, between medicine and educational research. In both the early vaccine non-experimental trial and in the Head Start evaluation, there is something we call a "confounding variable." In this case, family income.

As I said in the vaccine case, the higher income people were more likely to get the vaccine, but also more likely to get the disease and therefore the evaluation that didn't use random assignment was biased against finding the effect of the vaccine.

In the Westinghouse study of Head Start, the evaluation was

biased also against finding an effective Head Start because in this case the Head Start group kids were higher in levels of poverty which was associated with lower achievement.

So the power of experimentation, and this is a point Valerie made very clearly, in random assignment is to eliminate confounding variables. You see, we could match the kids -- we could have done a better study than the first one. We've done many better evaluations since the original Head Start evaluation -- we could match people on the basis of family background, making sure that the people we're comparing are the same with respect to income or other social indicators. But we can never be sure that we have matched on some of the relevant confounders. Variables that predict getting the treatment that are also related to the outcome are confounders. And with random assignment, we eliminate confounders, all confounders even the ones we haven't thought of, and that is the power of the experiment.

Now, this leads to a series of questions and answers that really form the basis of this paper and I will go through them rather straightforwardly, through them rather quickly here. I've got actually ten of them.

The first one is: Am I then saying that only studies that use random assignment are scientific? The answer is no, I'm not saying that.

First, a randomized trial is relevant only when there's a causal question on the table. There are many terrifically important questions for educational policy that are not causal.

For example, this seems so simple, but have high school graduation rates changed over the past ten years? Which kinds of kids in which kinds of cities and states and in which kinds of schools are at highest risk of dropping out? Tremendously important for policy to know the answer to that question. It is

not a causal question. We need a carefully designed survey to answer that question.

So, not all questions are causal. But, secondly, even when a question is causal, it may be impossible to do a randomized study. Another analogy with medicine: researchers have come to a strong consensus that smoking causes lung cancer, but we never had a clinical trial where we randomly assigned people to smoke two packs a day. Yet, we had a variety of scientific inquiry that led to a strong conclusion. We need to know how family conflict effects school achievement, but can you imagine the experiment that would test that causal hypothesis? (Laughter.)

Third, randomized experiments sometimes create artificial circumstances that limit the generalized ability of their findings. I won't go into detail, but sometimes you need corroborating evidence from studies in a natural setting that aren't randomized and across -- the randomized evidence might be crucial, but you need to supplement it to see whether a new program works in a less controlled setting.

The second questions is: Suppose we do have a causal questions, how do I then judge the scientific quality of the study that doesn't use random assignment? I guess, what I would say here is that in all of science at the heart of it is an obligation of the researcher to systematically and painstakingly alternative explanations for any finding of interest.

So, if I see a study over here where these kids had a new writing program and these kids didn't, and these kids, the kids in the writing program are doing better than the ones who didn? t, I don't just say, "That shows the writing program is good." I think about other explanations for why that might have happened and I evaluate them. It's harder to do when you don't have a randomized experiment, but it is still essential.

So, a scientist is expected to search for disconfirming evidence, and that's a crucial feature.

Even if we did a randomized experiment, let's say we did the writing study, we randomly assigned kids to do the writing program or not, we'd still need to develop alternative explanations for why the program worked. The experiment might tell us that the program works. But we want to go further to know what are the crucial ingredients because that may be very helpful to practitioners and policy.

So, even in the randomized context we need to search for explanations, alternative explanations, disconfirming evidence.

Moreover, randomized experiments are never perfectly implemented. So, people who drop out of the study, you'll have missing data in the two groups. We still have to worry about subtle or not so subtle biases.

So, what makes a causal comparative study then is not simply whether there was random assignment, but whether the investigators have effectively, critically evaluated competing explanations for what was found.

That leads to my third questions: Isn't it a little bit Pollyanna-ish to expect this scientist, this investigator to police me, let's say, to police myself and I'm a human being with biases and I'm supposed to evaluate all these things. Well, the key point here is the burden of objectivity does not fall entirely or even primarily on the shoulders of the individual investigator.

The role of the scientific community is key. It's a healthy scientific community who can -- and this relates to democracy, being able to freely evaluate alternative points of view, not feel that there's going to be some censorship.

The people who are committed to the principles I just mentioned who evaluate this, the process of objectivity really involves this group of people

engaging in this ongoing debate. Scientists, as was mentioned, are trained to be skeptical and that process can really work. What's really in the final analysis scientific is what the community of scientists says is scientific.

How am I doing for time?

MS. NEUMAN: You're doing okay.

MR. RAUDENBUSH: So, now, so far, if we have a causal question we'd like to do a randomized experiment, we may not be able to, if we can't, we'll do it as scientifically as we can, and then sometimes we don't have causal questions.

This kind of takes us back to a prior question: Is it really possible to do randomized experiments in education? I would argue, yes.

The Tennessee class size study, which by the way Frederick Mosteller called the most important educational study in decades. An amazing state-wide randomized experiment to evaluate the impact of large versus small classes. I'm sure you're going to hear about some more of them today actually a little bit later in the next session, if I don't talk too long.

Thomas Cook has done two randomized experimental evaluations of the James Comer whole school reform program. There have been many randomized experiments in schools on the effectiveness of drug prevention programs, not as many though on instruction which is interesting.

So certainly they can be done. The fifth question then is: How can we do them ethically? In the paper, I sketch some scenarios where we can very ethically, very practically, very feasibly do large scale experiments.

Often, what will be randomly assigned to treatments though will not be children. It may, in fact, be schools. Imagine a popular program, I mention "Success for All" simply because as an early literacy program, it's a program that has

-- there are over a thousand schools already in it. Many schools want to get into it, but it's expensive. So, a lot of people want to get it, but they don't get it. And also the people who run that program can only implement it in so many schools in any given year.

We could run an experiment where we asked people to sign up who want to do it, perhaps give it to them free or at a reduced cost and just say there's only one condition, we can't give it to you all at the same time. We're going to have a lottery that's going to determine who gets it first which is a very fair way of deciding who gets it first.

So, during that interim period where one group of schools has started to do the program and the others are still waiting, you have a randomized experiment, and a very ethically organized one. That's just one example. There are other ways.

We need to learn how to do this. People didn't think you could do it in medicine. Like I said, the Salk vaccine trial was incredible, the double blind experiment. We need to be able to make the argument and we need to learn how to do this stuff.

Number six. I mentioned that not all scientific questions in education are causal, and can I give you a few examples? I'm not going to give you too many. But I do want to mention that we may not have been doing such a good job in education of doing impact studies, causal comparative studies, what works. We need to do a lot more of that.

We've done a pretty good job of doing scientific surveys, though. Large scale, national longitudinal studies, tremendous amounts of learning have come out of those studies. And, I'm on the -- I'm going to toot the horn of -- the AERA National Science Foundation grants committee which has given out small amounts of

money to large numbers of young investigators. We have a report that shows hundreds of terrific scientific contributions coming out of that, but generally not of the strong causal character because it's really based in fact on survey research.

So, we have done pretty well there. I won't go into the examples in the interest of time, but there are lots of them.

Number seven: How are the best non-causal studies judged? There is this class. We can't just forget about the fact that a lot of the scientific research is not causal. So, we have a bunch of questions: How did we select the sample? Do they represent a population? How do we measure the key constructs? Is there an established reliability and validity to those constructs? Was the analysis done accurately? Were alternatively explanations painstakingly assessed?

Those are some principles. But, once again, the key point is in the final analysis it's scientific peer review that applies those principles in a case by case way to evaluate the credibility of the findings.

So, number eight. I've only mentioned quantitative research. Does qualitative research play a role? I would say, yes, without doubt. Because we need to not just test the impact of things out in the field, we need to do a lot more of that. We haven't done enough. But we have to have good things to take into the field. We have to have good ideas about how to teach math, how to teach reading. Those ideas come from up close, careful study of expert practitioners in real settings and how kids learn. So, we need that up close kind of research but see we've got to do a better job of connecting that research with field trials of what works, and that's what's really been missing.

Number nine. I ask: How do you combine insights? If I've said you have to have experiments and you have to have surveys and you have to have quality of research, how do you combine the insights from the different kinds of

inquiry? I hate to go back to a medical example, but it's a very telling one. It's the causal relationship between smoking and lung cancer.

As I said, you couldn't do an experiment to make people smoke two packs a day, but you could do a randomized experiment on animals. Strong causal inference, but generalized ability to humans? Then, we do good non-experiments, or quasi-experiments or at least comparisons between smokers and nonsmokers using the best possible survey methods and qualitative research.

Here the analogy is looking at lung tissue and finding out that the lung tissue of smokers is damaged in ways that we might think would be linked to cancer. You put them all together and the weight of evidence, the experimental evidence on the animals, the survey evidence on people and the lung tissue -- qualitative, put them together and you get a very compelling case.

We need to do that better, and that's going to require a very effective and active scientific community.

My tenth and final question is: Is there any danger here that we are going to be overselling the role of science in education? I think there is.

I've got a quote here from E. L. Thorndike who wrote the lead article in the founding edition of the Journal of Educational Psychology in 1910. I won't read the entire quote except to say that Thorndike felt that a scientific psychology was about to produce decisive evidence on virtually every practical question that arises in education. We know in retrospect that he was wrong. Unfortunately, by overselling what science can do, it led to a crisis of, you might say, rising expectations that couldn't be met. For a long time thereafter science in education fell into disarray.

The same thing happened in the '60s with scientific problem solving, the idea that we would have kind of a social engineering model. We'd try

programs, we'd evaluate them, we'd get feedback, the programs would get better and the great society was going to be born out of this sort of scientific and engineering model. That was an overselling. We couldn't really pull that off.

So, let's make sure that we have a balanced view this time. I am so excited that we have an opportunity to do it, to do it right without overselling it this time. I am delighted to have had the chance to be here because I think we're at a point in history where there seems to be for some reason a confluence of factors and the determination of people who have some power here who organized this, to really improve the quality of research in education and the link between science and education and practice.

Thank you very much.

(Applause.)

MS. NEUMAN: A wonderfully wise man.

I know that we went a little bit longer. What we'll do is, I think, take a break and then come back at quarter to, and then what we'll do is we'll combine the two discussion sessions, since I really do want time for questions.

Our next set will be more practical implications in terms of our programs.

Have a good break. There's coffee in the cafeteria, a good Starbucks across the street.

(Whereupon, the foregoing matter went off the record at 10:24 a.m. and went back on the record at 10:44 a.m.)

MS. NEUMAN: We're going to get started again.

Let me tell you that all the talks are going to be on the web, as well as in print. I know I forced people to rush through their presentations. The more complete presentation of each will be available to you immediately to you on

the web, and, then, in a little bit longer period, in print.

First, before we begin our sessions, I'm just delighted to introduce Linda Wilson. She's the deputy in OIIA, the Office of Intergovernmental and Interagency Affairs. Did I do that right?

MS. LINDA WILSON: Yes, exactly.

MS. NEUMAN: Good.

MS. WILSON: Hi, I just wanted to make a very quick notice. The department tomorrow is going to be releasing publicly a draft of the strategic plan. It will be on our website. It communicates the President's and Secretary's priorities for education over the next five years. It has very strong accountability, much like the No Child Left Behind Act, and it will guide our work here at the department.

It sets high expectations for us and it provides leadership to the nation's educational system. It's built on six strategic goals, which are create a culture of achievement, improve student achievement, develop safe school and strong character, transform education into an evidence based field, enhance the quality of and access to post secondary and adult education and establish management excellence.

The plan will not be nor should it be a trophy to hang on the wall. It's a living document that will guide the course of our work here through the next five years.

Secretary Paige is very committed to this. He has announced his intention to hold each department of education program, office and employee accountable for their responsibilities for implementing this plan.

The reason I am telling you this is because we would welcome your input to this process. As I said, it's going to be available on the web tomorrow.

Your comments we would need by 5:30 p.m. on Thursday, February 21st.

MS. NEUMAN: Thank you very much, Linda.

Now, we turn to implications: What are the implications of a scientific based research approach to our programs, so many of our programs that are going out to children?

I'm asking each of these presentations to be real brief because I really want to give you opportunity to ask questions and make comments.

The first presentation is by Russ Gersten. I have read so much of his work over the years. He's at the University of Oregon. He's done a lot of work on reading comprehension, teacher knowledge, and today what he's going to be talking about is the scientific based evidence and what that means for math education and achievement.

MR. RUSSELL GERSTEN: This is actually an easy topic to be brief on because there isn't a lot of scientific research in math. There's some. There's some promising directions, but it is a somewhat depressing topic.

There are two things going on. One, in elementary education there is no question that most teachers, even most parents, -- the reading is the big emphasis there compared to math. But it's not that simple. For other reasons, the math community of math educators at least for forty-plus years has looked at their role as reform, as change, as re-conceptualizing.

Therefore, there hasn't been this steady tradition. There are a few exceptions of really systematically using the methods that Valerie and others talked about earlier to build a knowledge base, but rather to study using the more qualitative methods: teachers understandings, kids understandings.

So, this is something that can change. There have always been little glimmerings of change. There's a slight increase in the amount, but overall the

math education community has been quite resistant to that, where let's say in the reading field there have always been at least two schools of thought, one in the experimental group.

But rather than just dealing with how little we know and getting us all depressed, I am going to give some highlights of some work we recently did actually for the state of Texas who was beginning a big initiative in the area of math, getting kids ready for algebra. So, it was basically, these kind of low achieving kids who got to middle school and just were weak in all areas of math. We tried to put together the scientific research, using the procedures we've heard about in terms of meta-analysis and all, in the area of math for low achieving kids. I did this with my colleagues Scott Baker and Dae Sik Lei.

I'm going to quickly go through the criteria, and they resonate with what we've been hearing about during the first session. We looked for studies that used random assignment. We did include the quasi-experiments, the ones that are kind of close, but they only were included if they had measures showed that the groups were comparable at the beginning. So, if they just used the school down the road, they were thrown out. They had to have at least one math performance measure, which sounds weird. But there were articles published in journals that either had teachers grades or students attitudes or certain interviews that we had no idea were they valid or reliable.

We found four categories. Notice the small number of studies we found on this. Now, we limited ourselves to low achieving students. These were students whose documentation was well below grade level, at least below the 35th percentile on some standardized measure.

But some of the things that worked, and again we don't have a lot of replications, but they were pretty decent studies, is that when kids and/or their

teachers get ongoing information, every two weeks, every four weeks, of where they are in math in terms of either the state standards or some framework, it invariably enhances performance.

This sounds kind of a little boring, it's not as romantic, there's so much of romantic work done in math. But the idea of having a system to know where kids are and what they really know, rather than saying this kid is struggling, this kid is struggling with fractions, manipulating fractions, more than one, with dividing fractions, with a sense of place value once you get into the hundreds. That information can be critical for low achieving kids, can be a life or death issue.

The second group we found, there was only six studies, is peer assisted learning. It's usually tutoring. This is something that could revolutionize practice. Invariably, when kids are partnered up, and it seems to be better if they're heterogeneous pairs, there's one stronger student and one weaker student and they switch off, achievement in math is always improved.

So, peers can be excellent tutors. I'm not talking here about cooperative groups of four, five, six kids. It's two. And if you see the difference in classrooms when there are two, it's very easy for the teacher to quickly monitor and get a sense of what's going one. Because kids are either working on stuff together, giving each other feedback, taking turns, or they're not. When it's a group of four or five, you're never quite sure what's this group discussing, these two kids look zoned out, but maybe they're finished.

So, the advantage of this, again we're not dealing with these profound things but with these kind of building blocks of improving practice and especially if this is based on the kind of data we were talking about can lead to reliable, replicated improvements in performance.

The one thing about the studies, and then we'll go on with the

finding, is that 60 percent of them used random assignment so they met the gold standard. Another third were this quasi-experimental group, so overall the small set we had were of good quality. And seven percent were partial -- they randomly assigned teachers and gave us some evidence that the groups were equal at the beginning which in the scheme of things is very, very good.

This is something that wasn't discussed so much earlier and is critical is did somebody come in and see were people doing what they're supposed to be doing? Because one of the key findings from the 1960s is sometimes these evaluations were done of people who were supposed to be doing science this way, or math this way, reading this way, but there was no evidence that they were really doing it. And, in fact, when people did drop-in site visits, they found they were not doing it.

So, two out of three studies did have an observer come in once or twice a week and make sure the thing was happening which sounds mundane and all but was a critical thing. So the quality indicators of the studies were good.

I'll go back to just kind of a quick summary, trying to speed this up. With the peer-assisted learning, the six studies consistently showed moderate effects -- and I'm not giving the exact numbers, but there's statistical ways to cut across called meta-analyis -- and that is an important finding.

When kids saw the data, and it was almost always on the computer, how they were doing, which skills they needed work on, whether they were making progress, these were moderately large, these were pretty large. This was especially true not so much for special education students but for that other that kind of at-risk group who are sometimes in Title I programs who sometimes need tutoring, that giving kids this kind of feedback seems invariably to help.

A very small number of studies on instruction. We broke them

two ways: explicit instruction, that includes both the very, very heavily tightly sequenced work that Carnan and some of his colleagues did in math which has everything sequenced exactly for kids and a beautiful array of examples, and some of these other approaches to teach kids problem solving strategies.

In both cases, and we only have a small set because we're looking kindergarten through eighth grade, but there is some evidence that providing this degree of explicitness to kids, showing them strategies, letting them take over and showing what they know is helpful.

This is hardly a revolutionary finding but it is important because there are many in the schools who do not advocate for such practice. This is invariably useful and when that's removed from children, especially the children below average, it tends to lower or decrease their achievement.

Contextualized instruction was our way to fit together very, very, very exciting ideas about the discussion teaching fractions and getting kids immersed in real world problems that involve measuring and fractions and equivalents. And the results? I put a question mark there. When we averaged them together -- and again we're only dealing with four studies -- it came out about zero.

So, basically, there is something there but how to get it into an effective package requires a lot of work.

This is an interesting thing. There were only two studies here that were done in inner city Philadelphia schools in terms of giving concrete feedback to parents on how kids are doing. These are low achieving kids and we're getting into the middle school years.

What the researchers found and they did two things. They set up the tutoring, was one thing they did, and then using this randomized idea for about half those kids and about half of the control group kids they also gave the parents

feedback when the kid was doing well.

This was their reasoning -- and this isn't the only approach in terms of communicating with parents -- that often by middle school when kids are D students and basic math, whatever it may be called, the lower track courses, in tending to get feedback it tends to be very negative. So, the teachers, if the kid was having problems, they gave that information for the peer tutoring session. But when the kid did well, they sent notes home, they called, now they could e-mail -- these studies were done a while ago -- and said you're kid is doing well you folks should celebrate this. Go walk up the mountain, a pizza party, whatever it is. So that the parents started to know the weeks, their daughter or son was doing well in math.

Now, that isn't a lot. I wish to say we had a hundred other findings. We don't. I just have a couple thoughts towards the future. Susan, if I could have a couple of minutes?

There are other lines of research that are not controlled intervention studies taking place in classrooms. I think we need hundreds more of those studies. Because as you see from this very small group of approximately 15 studies, we found some things that could be immediately useful for helping the below average, the at-risk kid in math.

But in terms of really conceptualizing and thinking about math, a couple of just my thoughts on what I envision is. As in the area of early reading about twenty years or so ago there was this insight and some beginning work on the phonological or phoneme awareness idea and how critical that was. Initially, it was very vague and no one quite knew what to do with it. There were some programs that seemed to have parts of it. It took a long time for that to solidify.

There's some very, very interesting work especially done by the late Robbie Case and Bob Siegler and others, in the beginnings of math. And, at

least in math, unlike years ago, we do have some measures that can predict. In kindergarten, we're doing some work in Eugene Research Institute in both Oregon and Texas at looking at predicting things by the end of kindergarten that will tell you which kids are likely to be at-risk. So you can start to screen and get a sense of stuff.

So, we do have at least a couple of measures that seem to validly predict and I know David Gehry at NIH is doing some work along this lines. So, we're maybe twenty years behind reading in this early intervention mode in terms of starting in kindergarten, starting in preschool, but we can move a lot faster now. We have the model of what succeeded in reading.

The other thing is we have this concept which is still elusive called "number sense." You'll see it around a lot. Nobody knows exactly what it is. It's sort of a sense of numbers, the way some kids just sort of take to it. You ask them, well, you know, here are six things, we want nine, how many more do you need? They'll just go "three." And, others will just go, "Well, you need some more."

But, it's just basically, the idea of both performing and understanding and doing and strategizing. We have his general notion. It seems a fascinating one. It seems a wonderful spur for a generation of new researches to do the kind of array of scientific methods. So, that's one huge area.

And I'm only going to do one other one. But this is something we've thought a lot about. One reason there's so little intervention research in education is people who've done it you leave totally exhausted. You're developing a new curriculum, you're training teachers, you're going in to see are they implementing it the right way. You're problem solving. You're going, oh, my god, why did we sequence the fourth week this way. You know, these things happen.

Then you're trying to develop valid and reliable measures. You

know, you do one or two of those. Then you say, well, maybe I'll do more, you know, literature reviews or correlational studies or descriptive case studies, because it is absolutely exhausting.

(Laughter.)

And you look at any discipline, and it's amazingly few people who have the endurance to do this.

But one system that the late Ann Brown developed is a very good one. What it calls for it says let's be honest. You can't just run in there and say this is a good way to teach math problem solving, where kids learn the stuff and then they practice in context. You need a while to do what she called "design experiments." To really go in and see what happens and collect data and not do the control groups and the randomization. You need one or two of those to get the thing working.

And they are not really just pilot studies. They are serious investigations of taking these phenomenal insights from cognitive psychology, from developmental psychology, but trying to put them into useable packages that there is some data to support.

Math is a long way from this. But this combination of doing the design experiments, but then not stopping there, to then test with the kind of controlled studies we were talking about before.

Those to me are the two at a national scope for future research. In terms of the last one, towards the future, I think because we're seeing such consistence sense that when the teachers or kids get ongoing data where kids are and what they need to learn once a month as opposed to once a year. It's a great way in October to say, you know, this kid doesn't know how to multiply fractions. So, he's in the 7^{th} grade, but let's get that under her belt, his belt, so we can move forward and this kid isn't going to get lost in pre-algebra. So, we need strategies and

measures to get this into practice.

The last thing is, as we look at what's going on in the field, we could do as twenty years ago Thomas Goode and Douglas Grouse did, which is look at what's happening in schools and try to link them to outcomes. Because we've got a huge array of measures in math, but we don't have a sense of which ones lead to better achievement or not.

So, those are my four thoughts towards the future and my sense of some pockets of knowledge we know for this average population.

(Applause.)

MS. NEUMAN: It's delightful to have Dr. Eunice Greer here today. She has done much work in the state of Illinois and been a director of reading as well as assessment. Today what she is going to be talking about is implications for scientific based evidence approach in reading.

DR. EUNICE GREER: Good morning.

It really is a very cool time to be working in reading.

Leave No Child Behind. No Child. It is a horribly devastating thing, and I'm not exaggerating, to be the seven or eight year old sitting in the room who can not read.

The next time you are in a classroom I want to challenge you to pick the 5 percent to 15 percent of the children in that room who will not learn to read and figure out how you're going to tell them that it's okay. How are you going to tell their parents? It's not okay. Leave no child behind.

Russell's right, we're fortunate in reading. We are beginning to build and see a converging body of evidence that tells us that we know something about successful strategies, successful elements that need to be taking place in early reading classrooms that will help ensure that all children learn to read.

We have a converging body of evidence that tells us that children need instruction in five areas: phonemic awareness, phonics, fluency, vocabulary and text comprehension.

Now, twenty years ago, when someone would say to us, well, how do you teach kids to read, we were left standing there with our hands in our pockets saying, well, a lot of different things work for different kids. We've come a long way since then. It's much more comfortable. I'm much more comfortable standing up here this morning, then I would have been fifteen years ago, saying, well, there's a lot of stuff that might work, and if one thing doesn't work, try something else.

Most of my comments today are drawn from the National Reading Panel Report that was delivered late in the year 2000. The panel sifted through over 100,000 studies and the sieve that they used to sift these studies through to identify the studies that met their criteria for inclusion in their analyses were the studies had to come from a refereed journal, be published in English. They had to focus on reading instruction for children pre-K through grade 12 and they had to use experimental or quasi-experimental research design with control groups or with multiple baseline methods.

Now, as Valerie alluded to earlier, if we had just gone for straight experimental design, there was not a lot there, and we still have a whole lot of work to do.

But as the panel looked at the studies that emerged from their sorting and as they read the results, findings began to converge around these five elements of early reading instruction.

What I want to do today quickly is take you through those five elements and talk briefly about some of the truths and some of the misconceptions.

Speaking with phonemic awareness. What is phonemic awareness? Well, it's the ability to notice and think about and work with the individual sounds in spoken words, not written words, in spoken words.

Before children learn to read, they need to know that words are made up of one or more sounds, and that you can take those apart and change them and that they make different words. They need to be able to work with speech sounds.

So, if they can do this, if they're phonemically aware, where are we? What do we know about phonemic awareness? Well, we know that we can teach it. There are systematic instructional practices that we can use to teach kids to become more phonemically aware. Children who are more phonemically aware are better at learning to read and to spell, and it also influences young children's comprehension.

Phonemic awareness in the classroom is noisy. It's not doing worksheets because it's working with sounds. So, if you go into a classroom and all these little five year olds have their heads down and those big logs in their hands that we call primary pencils, they're not working on phonemic awareness. They need to be making noise. It's most effective when teachers work with small groups of kids.

Now, let's look at the flip side. What are some of the misconceptions around phonemic awareness? Does it assure success as a reader? No, this is not an endpoint. There are a lot of other things that have to go on before we have a successful reader.

Is it the same thing as phonics? No, phonics we'll see in a minute. It is not the same thing as phonics. It's about spoken sounds.

It is just for at-risk readers? No, the research tells us that all kids benefit from being more phonemically aware.

Is it a perpetual element of K-3 instruction? Does it need to go on every day for four years? No, 18-20 hours for most kids. Now let me tell you, if you haven't been in a building in a while, kindergartners spend more time in the bathroom in a year than 18-20 hours. It's a finite thing that needs to go for kids.

Phonics. Phonics teaches kids the relationship between written language and sounds so that they can use it to read and to write words.

Kids who receive strong instruction in phonics are better at decoding and spelling, K-6.

Explicit, systematic instruction in phonics is better than sort of random or nonsystematic instruction or no instruction at all.

What do we mean by "systematic instruction?" It means that we teach children letter sounds and relationships and then we let them practice those on things that they're reading. We don't ask them to spend a lot of time reading things that they haven't learned to recognize the sounds.

So, if we're working on "B"s and "A"s and "T"s, we don't ask kids to read the word: can. We work on words like "bat" and "at." And, we give them practice using the tools that they are learning, so that they see the efficacy of those tools and they begin to see and discover the routineness and some of the patterns in our language. Phonics instruction is most effective when it's begun in kindergarten or first grade.

Now, some of the misconceptions. There's one best program. There isn't. When the panel looked at the research on various programs of phonics instruction, there really were no significant differences in the effectiveness of the programs that they looked at.

Phonics is just for kids who come from low SES backgrounds. No, that's not true. Phonics is of benefit to all kids.

Phonics instruction is effective when it's taught as a supplemental workbook activity? Here, again, no. This is not a workbook activity. This is an activity that involves repeated practice in applying phonic skills to reading and to writing, so that kids have an opportunity to write and read and see how this tool is working.

Here, again, it's not an entire reading program. It is not an end. It is a means to an end. We're working toward comprehension.

Fluency. Fluency is the most neglected skill or element of early reading instruction. When we say fluency, what we mean is rapid accurate reading with expression.

Now, when kids can read rapidly and accurately what this does is this frees up their little brains so that they can attend to what the text is about, they can attend to meaning.

Back in the '70s two gentlemen, LaBerge and Samuels, did some very nice research. They explained the notion of cognitive capacity. If you're spending all of your sort of brain energy sounding out words and trying to identify words, you have nothing left to attend to what the text is about.

So, we want to make kids as fluent as possible so that every ounce of capacity that they have can be put toward the outcome that we're looking for and that is their ability to comprehend.

Research tells us that repeated monitored oral reading practice can improve students fluency.

Now, the best strategy for developing fluency that we've seen coming out of the research is to give students many opportunity to read the same passage orally, and these need to be reasonably easy for the kids. They need to be at what we call their independent reading level, so they can read them with about 95

percent accuracy.

The best way to do this is to begin by providing kids with a fluid model of what this text sounds like, and then give them opportunities to practice reading it orally.

What are some of the misconceptions? Fluency is the same thing as authenticity. No, authenticity is just saying words right and fast. That's not reading with expression.

Fluency is a fixed accomplishment, you either fluent or you're not. No. You're fluency varies with the text and with the topic and with the conditions and the expectations for what you read. The same thing applies for young children.

Sustained silent reading improves fluency. We were a little bit surprised by this finding, but there's no evidence that sustained silent reading makes kids more fluent readers.

Now, there are a lot of hypotheses as to why this is the case. There is a lot of research that needs to be done, but sending kids off to read for thirty minutes by themselves and not holding them accountable and not asking them to practice is not associated with gains in fluency.

Let's go on: vocabulary. Vocabulary are the words you need to know to communicate. Oral vocabulary refers to the words that we use in speaking or that we recognize when we hear them. Reading vocabulary refers to the words that we recognize in print.

Students have an oral vocabulary. They have a reading vocabulary. Their oral vocabulary is typically much larger than their reading vocabulary. The larger a student's reading vocabulary, the easier it is for them to comprehend. The larger their oral vocabulary, the easier it is for them to comprehend and to read. Because when the come to a word they don't know, they

have a whole bank of words to try to match that up with and to associate it with. So, the more words they know, the more likely it is that they're going to experience success as readers.

Vocabulary needs to be taught directly and indirectly. Direct instruction in vocabulary is where the teacher introduces the word, discusses it, talks about it, lets kids write in sentences, work with it. Teachers can typically cover about 8-10 words a week in that method. That's not very many words when you think about how many new words a child is confronted with every week.

Kids learn most of their words indirectly, through conversation, through listening to adults read and talk and through reading on their own.

Misconceptions? Students can always rely on context to figure out unknown words. No. Beany Babies are ubiquitous. Could mean beautiful, could mean cheap, could mean really annoying.

(Laughter.)

Kids need other strategies to help them with unknown words. They need to know about dictionary skills and reference aids, and they need to know how to use those aids.

They need to know how to look at a word and its parts: prefixes, suffixes, roots. All of those strategies help them deal with unknown words.

Students either know a word or they don't. No. There are really about three levels of word knowing that we talk about. There are unknown words. There are words that you're acquainted with. You sort of know what they mean. "He went down to the cay to watch the boats." Well, I sort of know that's got something to do -- but I'm not sure.

And, then, there are established words that we really know well. They are our old friends. We know their multiple meanings. We know how they are

used. We know the affect that they convey. Those are words that are established in our vocabulary.

Finally, teachers need to teach new vocabulary directly. Obviously not, if a teacher can only cover 8-10 words a week, well, direct instruction of vocabulary words is not going to be the best and only way to go.

Finally, where are we going? Where is all of this headed for? Text comprehension, that's where we want to get kids. The other things are means to an end. They are contributing factors. But we always need to remember, our final goal is to get kids who are purposeful and active readers, and all five of the elements of early reading instruction play critical roles in contributing to kids getting there.

Truths about comprehension: good readers are purposeful and active when they read. They read for a purpose and they're always thinking and working through the text. Their brains are very active while they are reading.

There are six strategies that research has shown us that improve kids comprehension, six instructional strategies: teaching kids to monitor their comprehension; teaching them to use graphic and semantic organizers which are maps; sort of organizational pictures of the text content; being able to answer questions about what you've read; being able to generate questions about what you've read; being able to recognize the story structure --Is it narrative, Is it exposition, Is it chronological, Is it comparison and contrast?

All of those things are aids to being able to understand the text and being able to summarize a text.

Explicit teaching of these strategies, directly explaining the strategy, modeling it for the child, giving the child guides to practice with the strategy, giving kids repeated opportunities to apply and use the strategy. These are all effective techniques for teaching kids strategies to use when they are working through

text.

Misconceptions. It's best to wait until students have mastered the basics to teach comprehension. No, comprehension begins at the get go. We begin with listening and story comprehension, and as soon as they begin to read, we begin to teach them comprehension strategies. We don't wait until they're fluent.

Asking students questions about what they read is effective only as an assessment strategy. No. It is in fact an effective teaching strategy as well.

Finally, moving really, quickly, research implications. What are some next steps?

We don't have all the answers. We need to know a lot more. We need to encourage research that focuses on finding out more about the reading achievement and instructional needs of more diverse student populations, including students with disabilities.

We need research-based resources infused into the pre-service and in-service professional development systems around our country.

And, we can't forget principals.

Please, ladies and gentlemen: yes, teachers need to know how to teach reading, but those principals in those early elementary buildings need to know about early reading instruction. They really need to be effective leaders. If they are going to be effective leaders of reading instruction, they have to know it. They need their own professional development. They're not the same as teachers.

The field needs developmentally appropriate assessments that reflect what we know about early reading instruction. Teachers and principals need professional development around how to collect and use this data to inform instruction.

Finally, what can you do? Please, in everything that we think

about putting out there, we need to support and encourage teachers' use of research-based practices and research-based assessments.

We need to reinforce the need to teach all five elements of early reading instruction, and we need to remember that the goal is fluent readers.

I'll make a plea for consistency here. I talk to a lot of teachers. If any of you are standing up in front of a roomful of teachers, they are only going to see you once in their lives. Why should they trust you? They don't trust you. They're going to leave and go back and do what they did.

But, if we hit them again and again with the same message, it's a consistent message, it comes from all of our organizations, it comes from the Hill, the consistency proxies for trust, and they begin to listen to us and change and that's how we Leave No Child Behind. Thanks.

(Applause.)

MS. NEUMAN: You notice how Eunice's voice went up when she talked, "and principals."

(Laughter.)

I turn now to safe and drug-free schools. We're welcoming Judy Thorne from Westat.

MS. JUDY THORNE: Well, I have to say that drug prevention and violence prevention research is somewhere between the depressed scale of the mathematicians and the enthusiastic exalted scale of reading. I don't think we know as much about drug prevention as we do about reading. But, we know some things.

I want to talk about basically two strands of research in this field. The first is what's been going on in the Department of Education under the Safe and Drug Free Schools, or as it started out, The Drug Free Schools and Communities Act.

There is a progressive body of knowledge and this research primarily at least I find has been in the way of helping us to understand and know what's going on in schools in violence and drug prevention.

We started with a descriptive study that I had the pleasure of working on back in 1998 through '91 that looked at the initial implementation of the Act.

Then there was a longitudinal study that followed from that and used some of the information form the descriptive study to select a group of school districts that we then looked at longitudinally and drew relationships between the kinds of programs that they were implementing and the outcomes for students.

Some of the important findings from that study are going to crop up again in what I have to say. So let me briefly go over those.

One is that the differences between the groups, between very extensive and well implemented programs and the less extensive and less well implemented programs were small. They were significant but they were small.

Secondly, and this helps I think to explain the small differences, is that very few of the school districts and schools were implementing models that we were then coming to understand that there was a research base growing to support specific models of prevention education, and very few of those were being implemented in the schools for a number of reasons.

We also found that districts that had a full-time drug prevention coordinator rather than someone who shared that role with five or six other roles in their district, those districts with the full-time prevention coordinator had better outcomes, and programs that combined classroom and non-classroom activities had better outcomes.

Going on from there, there have been additional studies in the

department, one that focused on school violence, another that looked again at again at L.A. area school district activities.

There's a study going on right now of the quality and impact of safe and drug-free schools funded programs and the Middle Schools Coordinator Initiative where funding has been provided to actually have full-time coordinators focussed on middle school and research to find out if that's effective.

At the same time and sort of outside this realm of studies that focussed just on safe and drug-free schools, is a growing body of literature and findings to support specific ways of going about more often drug prevention education, but also violence prevention education, and I must say that they overlap a great deal because a lot of the risk factors in youth and in their communities are very similar.

So, based on a large number of studies, there have been a number of attempts to bring together a group of experts and sift scientifically through those studies to make recommendations about which appear to be the best models to use, mostly looking at classroom based curriculum in drug prevention.

So, we have several organizations or agencies. The Department of Education has had a panel to look at these and come up with exemplary and promising programs. The Center for Substance Abuse Prevention has done so. An independent organization called Drug Strategies has published a report on their rankings of prevention strategies. So, we have some specific curricula that can be recommended.

At the same time, others have been doing meta-analyses of these research studies and have isolated certain content that they believe is the most effective parts of these curricula and also instructional strategies that seem to be common to the most effective strategies and absent in the least effective strategies.

So, unlike the discussion that we just had in math and reading, I'm not going to go through the research and tell you what those strategies are, my main point here is that there are some established pieces of research and some knowledge of what ought to be happening in classrooms.

Are any of those programs perfect and absolutely, you know, doing away with drugs and violence among our youth? No, so we haven't reached the pinnacle of that kind of program development yet.

Nevertheless, when the Principles of Effectiveness came out in about 1998, and now are re-emphasized and expanded on in the No Child Left Behind legislation telling school districts and schools to implement research based programs, there are some places that they can turn to find out what those are and figure out what would be best used in their own schools and school districts.

Implementation issues. The other thing we know, especially from the studies we've done of what's going on, is that these research based programs are not widely implemented. We find very few districts and schools implementing research based programs.

A couple of studies, the study of L.A. activities done by the Department, and another one done by Chris Ringwald, Susan Annid and myself and others in North Carolina but looking at a national sample of schools and districts found that few are really looking at -- about 25 percent, I think, were implementing any of the recommended models. And almost everybody implement a whole number of curricula, not a single one.

But when you look at the content and delivery, things that have been isolated by meta-analysis, it's more encouraging. About 62 percent of schools reported that they were delivering the content that meta-analysis said was important, but not very many of them are using the teaching strategies that the meta-analyses say

is effective.

Now, why is this happening? Well, one is that there is not a big transfer of knowledge from the research community to the schools. Another is, I think, a lack of money to do this. I don't know how well the research supported curricula can be implemented on the amount of funding that they get from Safe and Drug Free Schools, which is about seven dollars per child, or could be reduced to around $3.50 if they decide to divert those funds for other purposes. So, they need additional funding if they are going to be doing those.

Another thing that I think is extremely important is pressure on time in class. The schools are under tremendous pressure to meet standards in academic areas. Unless they see and strongly believe in a link between the behavior and health of their students and those academic achievement areas, then it's really tough to make the pitch for a lot of time being spent in the classroom or in the school day on prevention activities.

Efforts to improve this situation. First of all, I've mentioned the Principles of Effectiveness have been out in the field for a few years and are strongly reinforced by the new legislation, and I think that that as it keeps being disseminated is an important piece.

The Middle Schools Coordinator Initiative is a way also of attempting to influence and improve the standard level of research based implementation in the schools. In terms of adding an additional person in the district who has time to really focus on these issues and figure out what strategies ought to be implemented and to implement them.

Obviously, we have a long ways to go. In continuing the research, we face a number of challenges. We've heard about the kinds of designs and methods that ought to be used in school based research. I believe that

experimental and quasi-experimental designs can be used. But they require very careful planning. They require large numbers of schools. They require enough time up front to really get your ducks in a row, get your entities selected. If it's going to be schools that you randomize, that can't be done sort of after the fact, after some schools have gotten funding to do something and go hunting around for maybe some comparable schools to compare them to. It takes a very concerted, planned effort of research.

I am definitely advocating that. That planned experimental or quasi-experimental designs be applied to specifically studying a targeted look at specific interventions implemented in the field.

I think this is what one of the earlier speakers was talking about in terms of field studies. Take the approaches that are research based or found in controlled studies to be effective, and look at them in a real setting in a number of school districts and schools at once.

Most of the research that we're basing all of our actions on was done in relatively small groups and much more controlled settings.

I am not talking about applying experimental design to a national evaluation of the entire Safe and Drug Free Schools program. I could do a very long presentation on why I think that, but I think I will move on.

What, two minutes? Oh, dear, this is very hard.

MS. NEUMAN: I'm sorry.

MS. THORNE: No, I completely understand. But, it's very hard to respond to, to try to pull all these issues in a field together and get them delivered.

One of the other challenges I wanted to mention though before I go on is the overburdening of schools. Where are all the schools to participate in all of the research that we've been talking about? There is not an infinite number of

schools out there. Many of them are already engaged in specific research activities. And, if they are not involved in a study of a particular intervention, they've been survey twelve times in the last year. It is tough to talk about this kind of research and then think about -- if you're in a school district or a school, you know how many times you've been asked lately to participate in studies. And you often have to turn them down because you just don't have the time available to do it.

Going on to the possibilities. It seems to me we are fortunate to have reached the point where we have some evidence to go on and some models to try out in a field based setting. And I think we can use experimental designs for some of these studies.

If, as I've said, we can have large enough samples, if we can have the time in advance to plan it and if we have strong support from the administrations of those schools. One of the challenges that I sort of skipped over is sort of the whole logic model of what is the intervention, how can you tell when it's well implemented, and how do you measure the outcomes? Measuring the outcomes in this area is tough. I mean, I hesitate to say this when we heard how depressing things were in math, but I don't think the challenges are quite equal across all of the fields in terms of research. You know, not being a math researcher I can blithely say that that's a lot easier.

(Laughter.)

You know I can somehow conceive of testing a kid's knowledge in math. Driving violence prevention, we're looking at stuff we can't even see. We're not supposed to see in the classroom. We want to know what those kids are doing when they're not in the classroom. How do we find that out? Well, probably the best way we've come up with so far besides urine tests is surveys. And surveys, well, all the schools are over-surveyed to start with, but secondly, we're facing the

Grassley Amendment which tells us not to survey students on sensitive behaviors, which illegal behaviors like drug use and violence are, unless we have explicit criminal signed consent. That just adds a further difficulty to the research there.

When we're looking at the possibilities, we should be looking at are the proven approaches affordable and effective in the real world, what new approaches are effective, and don't forget the non-classroom activities. Most of the research that I'm aware of at any rate deals with curriculum. And as I've said before in that longitudinal study, we've got a pretty good sense that the non-classroom activities were important as well. And by that I mean things that happen outside the classroom in terms of conflict resolution projects, student assistance programs, other kinds of things that happen in schools or around school time that is not necessarily classroom related.

And finally, I think our research responsibility is to continue to look at those targeted studies of approaches, but also to continue to monitor the implementation of research based programs in the school setting. So, I see a really important role in continuing descriptive research, looking at and talking with schools and school districts about the specific models they are implementing to find out if in fact that transfer is happening and to somehow help that happen.

Thank you.

(Applause.)

MS. NEUMAN: I know, it's so terrible. I'm rushing everybody, but I think you probably heard a startling statistic in that last presentation, which is 25 percent of all of the programs in Safe and Drug Free are research based. So, it doesn't seem as much an issue of money as much as a concern about dissemination and better dissemination of research based practices into those programs.

Finally, we are delighted to have Becki Herman who is a Senior

Research Analyst from AEIR talking about comprehensive school reform. Becki?

MS. BECKI HERMAN: Well, thank you very much for the opportunity to come here and talk with you about scientifically based research and the Comprehensive School Reform Demonstration Program.

I am going to cover three areas in my short time. Give you an overview of the research on CSR, and I won't delve too much into the actual findings that really focus on the quality of the methods in the research. And talk about what it means to apply the definition of scientifically based research to the Comprehensive School Reform Demonstration Program, CSRD. And also, to suggest some possible effects of using this view of research standards on the CSRD program.

First, I want to start off with a brief explanation of Comprehensive School Reform. What is Comprehensive School Reform? Comprehensive School Reform is a school level reform that's built around a unifying theme. It should be touching all grades and key subjects, English and math for starters, and it should touch all aspects of the school, and this is a key piece: instruction, curriculum, management, parent involvement, community involvement, school organization. There are a number of aspects of the school that need to be covered in Comprehensive School Reform.

Now, to facilitate Comprehensive School Reform many universities and private organizations have developed models that can be selected by schools and adopted by schools. But CSR is not just models. CSR can involve schools developing their own approach where they're thinking of how they're going to revise and revamp their instruction and curriculum and their management around this unifying theme, or if they chose to adopt a model, it might be adopting a model and working with other separate practices that they want implement in conjunction with this model, they all fall under this unifying theme.

Since 1997, the Department of Education has supported Comprehensive School Reform with a Comprehensive School Reform Demonstration program. It's not the only support, but it's one of the biggest.

I want to touch briefly on the state of the research. Much of the outcomes research focuses on models and so that's really what I'm going to focus on when I talk about the research but I want to remind you not to lose sight of the fact that models are only part of the story. There's a missing part of the story that's not necessarily being told because the research is a little weak there.

In the year 2000, the American Institutes for Research produced the Educators Guide to School Reform which profiled and reviewed the research on 24 of the most prominent CSR models in the country.

What we found was that there was limited research. We only found 130 outcome studies, and we set some limits for what we called an outcome study. It had to be focused on academic achievement and a few other criteria. And the new models have little to no research.

As part of the study, we rated the quality of each studies methodology. We used criteria such as what I have listed there under study methodology. We looked at the design. Was it random assignment? Was it causal, experimental? Did they use controls? What kind of construct, internal, external, validity evidence was there? What's the duration of the study? Was it longitudinal? What about the sample? The size of the sample, attrition, those sorts of issues. And measures? Independent and are they well-respected, high quality measures of outcomes? Independence of the researcher. Those are some of the areas that we looked at to rate the quality of the studies.

Of 130 outcome studies in 24 models, we found one study that met the gold standard which is true random assignment and also strong in all these

other dimensions of quality.

We found 61 studies that met the silver standard, that were quasi-experimental and strong in the other dimensions.

So, there's not a lot of gold standard, high quality, random assignment research. There is some research that uses quality experimental methods.

As Lisa and Valerie have pointed out before, the quality of the research base overall matters. It's not just the methodology used in the independent studies, but it's a replication of findings. It's that all of the research converges in a certain direction and points a way to a finding that can be useful to schools.

We found that there were very few models that had more than ten strong outcome studies and no models had absolutely consistent findings. There was always a school or a grade or a set of students that didn't do well with a certain approach. We were unable to come up with conclusive findings that said something worked well every single time.

But we were able to find that the bulk of the research, limited though it was, pushed in certain directions and that there were some models that seemed more consistent in producing strong student achievement outcomes.

It's important to look at the replication of findings, especially when you don't have a lot of gold standard studies, when you don't have a lot of random assignment studies because if you have hundreds of studies that are quasi-experimental study and no random assignment study, you might want to put some weight to those findings.

So, as I've said, I was focusing on research on CSRD models, there is very little CSR outcomes research that's not focused on models. OERI is currently sponsoring a set of studies that look at some of the issues that transcend models. They look at models and the study says well, but there are some issues that

are greater. For example, some of these studies are looking at what is the impact of comprehensiveness? Is the whole greater than the sum of the parts? Does a comprehensive reform work better than a set of discrete reforms within a school? Or some of the studies together are looking at the relative effectiveness of different approaches to CSR and some of the factors that help explain the variation.

In the last few years, there has been a marked increase in the amount of CSR research, including some random assignment experimental designs. The two Cook studies that studies that Steve Raudenbush mentioned earlier, a Success for All study that Steve Raudenbush described is actually one of the OERI funded studies where they're using random assignment and the issues that they're running into in conducting the study are too numerous to mention. But suffice it to say that they're committed to doing it and they've worked out a strategy for doing it, but there are real world issues with trying to do this.

So, now I've touched on some of the highlights of the state of the research on CSR, I'd like to turn to the circumstances under which the definition of scientifically based research should apply to CSRD. I'm borrowing from Baruch's chapter in an in-press book, Evidence Matters, for these five criteria for when you would apply the standard of -- for him he was saying random assignment studies, when you would use that standard.

The first criteria, the problem is serious. The second, the solution is unproven, other study designs will not provide satisfactory results, the results will inform policy decisions and the rights of participants can be protected.

Three of these criteria are easily met for CSRD: the problem is serious and the solution has not been unequivocally proven, although there's some evidence moving in some directions. And the results will probably inform such policy decisions.

However, the third criteria (that other studies will not provide satisfactory results) well, that depends on the question you're asking as almost every speaker today has said. If the research question is outcomes, does CSR improve student achievement, a causal question, yes, you'll get more defensible results using scientifically based research than using say case studies or some alternative design. If the question is what contributes to successful implementation, well, scientifically based research is not necessarily the only or the best strategy but certainly is part of the strategy for answering that question. But case studies can provide some very good information on what are issues with implementation and what are possible solutions.

The final criteria for applying the standards of scientifically based research to CSRD, is that the rights of participants can be protected. In this high stakes, outcome oriented environment for reforming schools that's a difficult criterion to meet. It's hard to ask a school to maintain a comprehensive school approach that does not seem to be working when they are under incredible pressure to produce results quickly for the duration of the study that you need to conduct. The study needs to be more than a few minutes.

(Laughter.)

It's also difficult -- and this is a problem with some of the CSR studies that are trying to use random assignment, there's the problem of getting and maintaining adequate comparisons. If you use random assignment, how do you guarantee that there's no slippage that they don't go ahead and adopt either exactly the condition you were testing or a competing condition, but, in other words, somehow tainting your comparison?

It's difficult to ask schools to either maintain or to not use a Comprehensive School Reform approach for the duration of a study, but there are

ways of doing it.

In situations where you're looking at outcomes and you're looking for causal effects and where you're able to protect the rights of participants, then it may be appropriate to apply the standards based research to Comprehensive School Reform Demonstration Programs.

CSRD in the No Child Left Behind legislation has eleven components. Only two of these components are explicitly tied to scientifically based research in the legislation. The first component which is "proven methods and strategies are based on scientifically based research" means the strategy for instruction should have some evidence using scientifically based research.

Then there are a series of components that talk about, say, professional development, measurable goals and benchmarks, that the design is comprehensive, which are less testable within experimental design. They are more about the development and the implementation and they are different sorts of issues.

But, the final component "that the CSR program results in significant improvements in academic achievement," the idea that the practices that you're using in your CSR program work and they work as a set collectively. That idea is also held to the standard of requiring evidence from scientifically based research or other evidence of effects.

I was talking to a few people before starting and some said that they were curious about what I was going to say and I said one of the first things I want to say is I'm not a soothsayer. I can't tell you how this new definition of scientifically based research will effect the program. But, I can make some suggestions of possible effects and I'd be interested to see what actually pans out.

One of the possible effects, focusing on the first component of the CSRD, the expectation that CSR programs use proven practices, one of the effects

may be the burden on the schools.

If you have a CSR program that includes a set of practices, you might have a practice like parent involvement. You might have a set of practices around curriculum. You might have a set of practices around instruction and a set of practices around management.

All these practices need to be proven. Somebody needs to go out there and do the research on them. There's no single source that says this is the best way to go about instruction or this is the only effective curriculum. So, a school that's thinking about adopting CSR needs to be able to investigate all these various areas of research and that's a huge burden.

That's a burden that can be eased with a lot of resources and I know that there's been mention already of the What Works Clearinghouse which will hopefully be able to provide some support for schools in this area. There are organizations, the Department of Education is not the least, that provide a lot of information to help schools look at the research. But, it's still very modest. That might deter some schools that are considering applying for CSRD grants if they're expected to look at all of these aspects.

If a school is considering adopting a model, there might be a positive effect of this new definition of scientifically based research that focuses on the practices. Schools will be looking at the practices within the model, not just the model. They might be able to see whether there is evidence for all of the practices, the curriculum practice, the instruction practices, the management practices, to see whether they think that this is the right approach for them and that there's evidence that this will work for them.

It might also cause them to question whether the model itself if comprehensive, whether there might not be some practices that are not part of the

model, say parent involvement, that they might want to investigate themselves.

Further, it might encourage schools to think about developing their own approach to comprehensive school reform that is inclusive of a larger series of practices.

So, this focus on finding effective practices may really cause them to rethink how they are using models and what practices they would like to be using in their reform approach.

A second positive effect of this definition of standards based research is the possibility that it might encourage schools to be critical consumers of research for them to look at whether something works. That is, provided, that, as I said, they have the resources to help them collect the research and have the resources to help them understand and interpret the research.

A third possible effect of the research standard is a possibly detrimental effect on externally developed CSR models which at this point is one of the most prominent subsets of Comprehensive School Reform.

There are a lot of different models, some that are more mature. They are in a lot of schools and have a strong research base. And then there are some that are smaller. They're newer. They aren't in a lot of schools. There's not a lot of evidence at this point.

With this kind of selection, schools can find a good fit for their own situation. They can find models developed around a theme that works for them. They can find a model that has a series of, a set of effective practices that they believe are right for their own strengths and weaknesses.

But new models may be strongly effected by the requirement for scientifically based research. If they are in few schools and they have not had time to develop a strong research base, this might prune the field. If you hold new models to

the same standards it might foreclose the development of approaches, so that you only have one or two big approaches that are mature for schools to be able to turn to.

It might be appropriate to think about a schedule of evaluations where you hold a different standard to the more mature models than to the newer models or the practices that comprise the models, or to support newer demonstration approaches differently from the more mature models in some way.

Finally, for all of this to work, for the research to be meaningful to practitioners, it's important to be able to build a bridge from the research to schools. I think I've mentioned this several times, researchers are trying to make decisions and they're held to the requirement that these decisions need to have some scientific evidence. So, one of the biggest movements I could see is providing more support for helping schools access the research and for helping them understand and discern between the various levels of research and quality of research.

(Applause.)

MS. NEUMAN: Well, as I look at the clock, I realize that all this prepared discussion time just has ended actually to be blunt.

I thought this was great evidence that the topic of scientific based evidence is truly a fascinating one. I was fascinated to see how many of you all stayed throughout the discussions, as well as the wonderful papers. I had read every one of these papers prior to today, and yet, I found the delivery of those papers still fascinating. The issues you raise are just really important.

We will be thinking about that as we give guidance throughout our various programs.

I'm sure people are willing to stick around a little bit after.

Again, I want to thank all for these wonderful presentations today.

They will be up on the web and please feel free to contact me or these wonderful speakers.

Again, thank you for coming.

(Applause.)

(Whereupon, the above-entitled matter was concluded at 11:58 a.m.)