

**RESEARCH IN EDUCATION:
ON THE LEADING EDGE OF SCHOOL IMPROVEMENT?**

**A PUBLIC POLICY FORUM PRESENTED BY
NATIONAL EDUCATION KNOWLEDGE INDUSTRY ASSOCIATION
PROGRESSIVE POLICY INSTITUTE
EDUCATION QUALITY INSTITUTE**

**WITH SUPPORT FROM
ACADEMY FOR EDUCATIONAL DEVELOPMENT
THE KNOWLEDGEWORKS FOUNDATION**

TUESDAY, MARCH 26, 2002

**ACADEMY FOR EDUCATIONAL DEVELOPMENT
1825 CONNECTICUT AVENUE, NW, 8TH FLOOR
WASHINGTON, D.C.**

**12:30 - 1:45 - LUNCHEON AND KEYNOTE ADDRESS
INFORMING EDUCATION PRACTICES AND POLICIES THROUGH
SCIENTIFIC RESEARCH**

1:45 - 2:00 PM - CLOSING REMARKS

*Transcript by:
Federal News Service
Washington, D.C.*

MR. FLEISCHMAN: Let's get started again. I have the great pleasure of introducing Dr. Reid Lyon, and doing this very briefly for a number of reasons.

First of all, his full biographical information is included in your packets if you want to see his multitudinous accomplishments.

Most of you also probably know Dr. Lyon from his work and, therefore, he doesn't need a long introduction.

Thirdly, he is going to take about 45 minutes. So the sooner I get off, the sooner he gets going and gets to provide you with what would be, I think, a very interesting presentation on the guiding principles of scientific research. Dr. Lyon is a research psychologist, and the chief of the child development and behavioral branch within the National Institute of Child Health and Human Development, NICHD, at the National Institutes of Health, where he is responsible for the direction, development and management of research programs in developmental psychology, cognitive neuroscience, behavioral pediatrics, reading and human learning and learning disorders.

He has been at this full-time at least for 10 years now, and I really do think that most of you in the audience know Dr. Lyon pretty well, and I will step aside and give him an opportunity to give his presentation.

So, thank you very much.

(Applause.)

MR. LYON: Good afternoon. I'm delighted to be here. I am following a stellar number of individuals who have blessed you with knowledge throughout this morning, including a number of researchers who covered much of what I may cover, although I'll try to embellish it. Lisa Towne, I heard, did a wonderful job on the new NRC report which has to be required reading for all of us who work hard in this profession. It's one of the best, clear, most elegantly written NRC reports I've seen in quite a while, and it's informative, its examples are outstanding, and I think it's going to help us out a great deal. Thank you very much, Lisa, and Dick Shavasin (sp), for bringing us that.

But, given that there are quite a few people that know what I'm talking about, I kind of feel like Zsa Zsa Gabor's seventh husband standing up here. I know what's expected. I just don't know how to make it interesting.

Before I start with some of the work that I'll talk about, and the way I'm going to try to do this is run through in more specific detail some of the things Lisa talked about, but I would like to weave into some of the criteria or conditions that we look at when we judge quality of research, and I would like to provide some examples from a number of studies that NIH and OEOI are

collaborating in to give you some practical footing for some of this stuff.

I would like to read something to you. The history of profession has never been a particularly attractive subject in professional education, and one reason for this is that it is so unbelievably deplorable a story. [Momentary technical difficulties.] It is hard to conceive of a less scientific enterprise among human endeavors. Virtually anything that could be thought up for treatment was tried out at one time or another, and once tried lasted for decades if not centuries before being given up. It was, in retrospect, the most frivolous and irresponsible kind of human experimentation based upon nothing but trial and error, and usually resulting in precisely that sequence.

The profession described in what I just read is not education, it's medicine. And this quote was given by Louis Thomas, former president of Sloan-Kettering, as he was bemoaning the state of science and state of art in medicine at the turn of the century, where, indeed, most of the things that practitioners did was anecdotal and based upon anecdotal experiences, idiosyncratic training, and primarily reflected what the internship experiences were. You could come out of Harvard with a medical degree or out of the University of Maryland with a medical degree and have absolutely different conceptions about biology, chemistry and physiology. And, indeed, it drove idiosyncratic practice. And the question, of course, given that a lot of people were not faring well given that kind of capricious practice of medicine, was, is this the way that it has to be. That is, is there a body of knowledge, a core body of knowledge, vis-à-vis biology and chemistry, physiology and the life sciences where all of our physicians to be could be provided a foundational set of concepts, and then specialize beyond that. And, indeed, today, after many years, obviously of honing that system, you can receive a medical degree from Harvard or the University of Maryland, and you are responsible for the same core knowledge no matter where you go, and that is assessed by different boarding structures, and so forth.

That is in no way where we are in education. You can literally come out of any college of education on the practitioner's side, or on the doctoral side, the research side, holding entirely different conceptualizations of what constitutes evidence, what constitutes instructional practice, what constitutes the assumptions upon which we base practice, what constitutes a canon of knowledge. It is in a sense tough to look at, and I don't mean to be derogatory or overly negative. I think we're making some headway, but there is no way, to be honest, we are ever going to stem the tide of this epidemic of underachievement of our children in this country unless we start to bite the bullet and get our stuff together.

I've known many of you in this room for upwards to 20 years. Many of you have been in the government, or in policy, or in practice for two decades, and we've revisited, and revisited many of the concepts I'll be talking to you about today ad nauseam without a bite.

I will not say this in a partisan way. I work for the president of the United States, as does all of the governmental employees in this room, and we are at a nexus, we are at an entirely different time than I've ever seen it. And whether or not it makes a difference, this is a gentleman who came to this job and gave us a couple of significant tasks to carry out.

One, staff these educational agencies with scientists who know both practice and the basic work, make sure they're outstanding managers, make sure you collaborate, because I want to see every initiative that we bring forth to help our kids, that you at NIH are tied to us, Whitehurst at OELI, are tied to Susan Newman at OESC, are tied to Ray Horn at HHS, are tied to Bob Pasternak at OCRS. Do not be carrying out splintered, fragmented initiatives, either on the research or programmatic side. And, indeed, that's now in place. Dr. Pasternak is standing behind there, and you can see we're linked at the hip, although I dress better.

This is an individual who said, we can no longer afford to leave our children to the mercy of malpractice, educational malpractice. Do in fact we have a corpus of information that is available in certain content areas that all of our teachers should hold, given that there's enough converging evidence to support its instruction to undergraduate and graduate teachers? Is that corpus out there? Why is it, in fact, that our teachers are leaving colleges of education with entirely different backgrounds, and approach education entirely different ways?

Why is it if you tell me that, in fact, we have that body of information in some areas of education and instruction that that's occurring? Why the translational difficulties? Tough stuff, tough work.

What I'm going to try to do is review what we know in one area for you, in reading. I'm going to intersperse that with a number of concepts that guide the conception of knowledge and evidence and scientific knowing. And it should be broad enough that everybody can fit what they do into it if what they fit into it is of, in fact, high quality, quality is critical. Lisa Keegan pointed out earlier, everybody needs to be under the tent working on all of this, but there is no reason to continue shoddy, mediocre work, it will not get us anywhere. Bad ideas won't get us anywhere. We have to be able to test those objectively and begin to move what we do with children to a higher plane. This is not about adults. This is not about ego, it's not about career, it's not about someone holding onto some silly ass ideas that have been out there perpetuating nonsense for years. You've got to let it go.

So, how do we let it go in a fair, objective way? How do we start to use what we know and discard what is harmful, or not useful? These are going to be busy slides, unfortunately, but I'm just going to run through them fairly quickly. But when we're talking about what constitutes scientific research, Lisa took you through a number of these general points, but just let me repeat them to provide the context.

Obviously, when we're trying to understand complex phenomena, we've got to reason extraordinarily cleverly. We've got to have some body of context, some theory, conceptual framework, if you will, that guides how we pose the questions. You just can't be doing research coming out of your head in an interesting way. We have to stand on the back of a solid body of integrated theory or conceptual information, because that's what's going to give us the grist for asking the questions.

We've got to learn how to be objective, and in science I think we are fairly objective. But as someone pointed out on the last panel, everything is biased. In fact, there's a whole school of

thought that says, because everything is biased, the way people think that is, then evidence and scientific data are in the eye of the beholder. That is, everybody's interpretation is as good as anybody else's. Right? Nonsense. The designs that we employ that drive from the theories or designs that should be able to pull the bias out when we are using control or contrast groups, or sampling paradigms, or whatever it is, not to be technical, but one of the major reasons we do the things we do in designing experiments is to unpack the bias and get it out of the way, and make the interpretations as strong as we can.

We know that no one study in education will ever, ever solve any problem. It is too messy, it's too complex, and we've got to be able to build, albeit in a plodding, circuitous, self-revising manner, a core hunk of information that has umph to it converging evidence. It takes a while. It takes a long time.

One of the things that, if you look at our federal research agencies that has been sorely lacking is that we have had fragmented, short-term efforts that simply did not provide the research community the infrastructure it needed to address questions that require longevity. Kids are moving targets, you don't study kids at any one point in time and figure out how best they respond to instruction, or figure out what works best. How do you do that? I mean, if you develop a certain instructional program and apply it to a third grader without knowing where that third grader has been, how do you know if it's the instruction that bumps the kid? Is it the instruction plus previous knowledge, and so forth and so on?

There is room for short-term isolated studies, there is tremendous room for basic science work. But when we are trying to figure out how kids learn, what it takes to learn, why they have difficulties, and what we can do about it, we've got to think strongly about providing the funding and the support and the leadership to create the designs that helps us watch kids move over time. We need to walk with those children as they move through life and their challenges in the classroom.

And a higher education person brought up earlier that that is anathema to promotion and tenure. In other words, if you are going to do those kinds of long-term, prospective, longitudinal, epidemiological, clinical trial kinds of things, it's going to take you all the way up until your tenure date before you get a publication. Well, why in the world would higher education be providing life-long guarantees when these short-term studies suck? Many of them do. They're not informative. You're counting papers. It's ridiculous. Maybe it's okay in history, maybe it's okay in philosophy, but it's not okay when we have to generate a body of knowledge that has to get to children, the human lives of children, not their philosophical underpinnings, not their political leanings, how they learn in school.

We start out, hopefully, with some good questions. That's what drives everything. And the quote up there is from Dr. Einstein as he said, the formulation of a problem is often more essential than its solution, which may be merely a matter of mathematical or experimental skill. To raise new questions, new possibilities, to regard old questions from a new angle requires creative imagination and marks real advance in science.

Not only are the questions critical, but we have to ask them in a way that they can be addressed, that they can be falsified, that they can be overturned if, in fact, they're not correct.

I'm going to try to weave in now some questions that have been asked since 1960-something, at least the first one has, to try to make this point. We've been working at NIH, and now we're collaborating with OEOI and OCRS on these questions for kids all up and down the spectrum from birth onward, but one of the things that we were interested in long before I ever got to NIH was the first question. How do children learn to read? In other words, what goes into it? That's kind of a broad, overarching kind of question. But what are the skills, the abilities, the interactions in the classroom, the environments, the properties of those environments, the genetics, the physiology that derives product by being able to handle, what goes into that? How does that happen?

And, of course, as you ask about the interactions, the skills, and the abilities, and the genetics, and the physiology that gives rise to all kinds of very specific questions about each of those major areas. But how do kids learn to read?

And we have been addressing that question since 1964, and we have a pretty good idea what it takes. Science is going to revise what we know today. It isn't going to be static. We're going to add and deflate some of the things in the concepts that we've learned. But there is enough in that first question, there is enough grist that has derived from the science in that first question to clearly begin to use the results of those studies in both federal legislation, and in teacher preparation.

You have in front of you H.R. 1, the new ESEA bill with components titled Reading First that builds on the body of knowledge developed through the NIH Reading Research Network, which is the largest in the world, and now collaborating with OEOI, NSF, and other agencies to continue this work. So we do know what it takes.

I'll take you through that in a little bit, but someone asked the question in the last panel, how in the world do you translate knowledge, even if you have converging evidence from the research community? How do you get it into the classroom? Well, teachers have to have an idea of how to answer that first question, and they don't in the main. I just came from a talk this morning with ostensibly the premier teacher preparation institutions in the country, making it here was by getaway car. People do not want to hear that we've got some work to do. They don't. And I'll talk a little bit about that as we go along. But, if our teachers understood, for example, what it takes to learn to read -- I think Mike Cohen was making an interesting point before that no one instructional strategy, no one program, no one approach is ever going to be equally efficacious, equally beneficial for all kids. No way. There is no magic bullet out there. We have never, in all the clinical trials we've been doing on reading instruction found a program that covered all kids that came in front of that program.

So what happens when you're only reaching 60 percent of children with the program? Then the teacher has to be able to modify, adjust, influence the representations that that program is giving, alter those in some way so that that program and its contents connect with the internal

schema, what the person has in their heads so they get it. But how can teachers do that if they don't know what it takes to learn to read?

In looking at that first question, it just wasn't that question, obviously. There were theoretical foundations that talked about reading occurring as a function of visual perceptual issues, of reading being parasitic on language riding on the back of semantics or syntax or phrenology, of reading having to do with theoretical ideas about the inner sensory vestibular system, of reading having to do with the amount of protein in hair, believe me, of reading having to do with a theoretical or a conceptual idea that eye tracking made a difference, some of these may resonate with the older folks here. There are theoretical positions about children learning to read as a function of developing neural structures from the brain stem forward as in philology recapitulates ontogeny. I don't know if you remember that one. But theory drove at least questions, and we have to base everything on theory. One of the competing theories that has been in place for a while that is actually getting pretty good support is that language undergirds reading. It is much more of a linguistic process than its competing alternatives, a visual process, or a vestibular process. And I'll take you through very quickly the converging evidence we have on how language drives the infrastructure, the skeleton for written language to hold on to.

But, at the same time, we were testing theories that learning how to read was natural, that is, if you watch little kids, you will notice that they listen and speak in a relatively effortless manner vis-a-vis instruction from parents, that is, parents don't have to directly talk to kids about adjectives, pronouns or adverbs, they talk back to kids, they reinforce the language, they expand it, and so forth, and little kids come along and do okay without direct instruction. And that, indeed, is true. Oral language is, indeed, natural. If we had that as a theory that guided questions, we'd find, as we have, that you can grow up in Tokyo or Toledo, you can grow up in Boston or Baghdad, and, indeed, you are going to come out and land on the planet and begin to produce sounds in exactly the same sequence no matter where the heck you're from. If you go to Tokyo, you're going to find out that like in Boston kids will coo and babble use singular, holifastic speech, and connect stuff together in an open context, and then extend syntax, and they're going to be fine. That happens everywhere without doing any teaching. It's an interesting, fascinating thing. It's genetically driven; it's protoplasmically wired; it's hard-wired.

But the theory, the assumption was that so as goes oral language, so goes reading. And that assumption, that theory, was never tested. It was just put into place. There was no scientific test of that theory. It, indeed, was based upon some observational capability that, to me, reflects low IQ, be that as it may.

If, in fact, reading is natural, why are there so many planets -- excuse me, why are there so many cultures on this planet that have very strong oral language cultures with absolutely no written language. And why is there in this country that is fairly literate such a high rate of kids who can't read. I mean, just at that.

So feeling your assumptions can stand alone without test, as in that particular assumption, or set of theories, based upon Chompskian work, stolen from Chompsky, actually, and used for other means, not tested, and we've got a lot of casualties out there who are probably in their teens

now not able to read as a function of that particular set of assumptions driving the instruction, dispedagogia, malpractice. That's what it constitutes, because it never was tested.

Here's something that I think confuses a lot of people, gets in the way of understanding and makes it difficult for the press, the lay public, practitioners to understand what we mean when we're talking about scientific quality. You hear time and time again that good research is experimental, good research is reflected in clinical trials. And you'll hear that qualitative research is not good research. And as education goes, as usual, people are now polarizing those who do quantitative research and those who do qualitative research, have you heard that one?

But, the problem is, it's the wrong question. The selection and application of the research design that has to take place for something to be scientific relies on two major principles. Number one, the design and methodology is appropriate for the question asked. Number two, the application of the appropriate design is done with high quality, is done well. If you are asking descriptive questions, or epidemiological questions, or if you're asking how people get in trouble with AIDS, you don't do clinical trials, you don't do randomized experiments, at least initially, for those questions. You have to bring to bear descriptive mechanisms. If you're asking about what works, then you have to start to look at cause and effect, does X cause Y. You have to tear apart everything to make sure that whatever is causing Y is X, rather than something in between X and Y, right.

So those are the major -- I mean, just to boil everything down when we're talking about scientific quality, it really falls on a couple of critical points, that is the methodology and design that we're applying to the problem fits the nature of the question, and that that application is of outstanding rigor, and intelligence.

Everything kind of hangs together. At the NIH we're involved in both internal research, that is we have our own research labs, as you know, on campus, and we fund a tremendous amount of research, biomedical research, behavioral research, cultural research, and so forth and so on. The peer review system is one of the oldest in the country. It's peer review, it's not perfect review. But, what it brings to this honing of our scientific knowledge is a capability of always looking at applications, or our own internal research in terms of these things coming up behind me.

That is, we always have to link together the question with the design, with the measures, with the theory, everything has to be tightly linked together. And believe me, at NIH, and even with our own internal labs, sometimes we have great theories, which have spawned pretty good questions about the universe, or how kids learn, or how brains work, and the design is totally mismatched with the type of question. And the peer review system then doesn't say, well, you dummy. The peer review system says, the linkage between your theory and your hypothesis, and the design that you're applying to test those hypothesis and analytic strategies are mismatched, and can we help you with that, here's some suggestions. Peer review, providing an outstanding, recursive, educational role, not a punitive role.

Now, after explaining that to the president, how it worked in response to his question,

how does NIH peer review work, how do our peer review systems at the department work, there are big differences he said, why can't that type of peer review also help our local schools, and our states? And I said, will, I think maybe it probably can, but we're going to have to be very patient, we're going to have to bring a tremendous amount of capacity to that task, we're going to have to make sure people understand these evidentiary issues you and I are talking about today.

But the point was, let's say that Reading First money, to speak practically, is ready to go out to 50 states, but it can only go out to the 50 states contingent upon meeting the criteria set in the legislation, very basic criteria. Those criteria in the legislation say clearly that if states and local communities want to use federal money to purchase programs, or to train teachers, those programs and the teacher preparation interactions must be based upon converging evidence of effectiveness. Simple. So the question is, do we have enough converging evidence of effectiveness in these areas where the money will flow? Yes, we probably do. Do the local and state enterprises know it? Maybe not, they don't, so we'll have to get that to them. But there is enough known about that to tie federal money to it, first time it's ever happened.

And the reason we're doing that is because this new law is not only a law that administers money, it's a law to begin to help a culture shift from anecdote, from belief system, from philosophical driven orientations, and so forth. It's a law that ties money to making sure that what is applied with our nation's children is, in fact, that which has been tested before, with those types of kids, and found to show effectiveness. And it is a law that says if you want to prepare teachers or provide continuing education for teachers, then indeed the money will flow when the application clearly indicates that what you're doing has been found to be effective with teachers of reading.

And we'll find clearly that the applications will come in on these Reading First initiatives, and they, indeed, will hold to the language of the legislation. We know the communities will hire grant writers up the wazoo. We know that they can use the language inherent in the RFF. We know from the Reading Excellence Act, which also had scientific criteria embedded in it, that once that money went out it in no way reflected in many cases what was written.

So now we come back to an NIH model, and we begin to look at the paper products, providing a plan to teach kids and train teachers, that goes through strong peer review. That's what's going to drive all this reading first stuff, not a list of programs, peer review, taking into account all of these features we're talking about right now, and what Lisa talked about, and I think a lot of the other panelists talked about. Peer review, that's what's going to let Minnesota, or Kentucky, I'd better not say states, whomever receives money in this formula grant initiative, which is a formula, but it's a contingent formula. State and local districts will not receive the money until their actions and interactions with kids reflect they know what they're doing. It just isn't going to happen.

And you know we're going to get all kinds of calls from Senator this and Congressman that. But, the fact of the matter is, all of these criteria that have been placed in Reading First in HL-1 were definitely and genuinely bipartisanly driven. What was interesting is that it took a while for our lawmakers to really understand the nature of what had gone on before, and what we

could do now. It was surprising to many of them that the amount of converging evidence on certain things that worked was as strong as it is, and it surprised them to understand that you do not, or should not employ that which does not work, because it's just not a lousy bet you lost, it's a human life.

So it took a while for them to get that. And many of our most senior leaders on the Hill, no names mentioned, stood back and said, and we've been supporting things that moved this process to nowhere, under the assumption that everything and anything should go under the assumption that education is different from other kinds of sciences. I think Lisa probably pointed out, in no way is education research different from any other kind of scientific research. I'll go through that in a minute.

But knowledge is going to come from using good common sense, and good common sense says, you have an idea about what works, or you have an idea about the nature of the universe, and that spawns some questions. And you've got to be very careful when you ask a question that what you are applying to address the question is capable of addressing it with high quality, linking everything together. We have been funding research in the federal government that doesn't even meet that particular linkage strategy ever since some of us have been alive.

The bottom part of this slide is the critical one, and I hope gets to the nub of some of the issues that you all may ask about or be confused about. To reiterate, the type of research being done, whether it's experimental research, ethnographic research, cultural studies done on a case level, correlational research, regression analyses, experimental trials, hierarchical growth models, what we are using as tools, design wise and analytically speaking, do not determine the scientific power of a study. The scientific power of a study is, again, predicated on whether you've asked good questions, that can be falsified, that we can test counter proposals to what the question is, or what the hypothesis is, and that we've applied appropriate methodology with clean rigor. And we can stand threats to internal validity, reliability problems, and external validity. Does this stuff generalize?

So when you hear these debates about quantitative, qualitative, and so forth and so on, we would do this country so much good if we could simply say, please stop asking that, please stop doing phonics-whole language, please stop doing quantitative, qualitative. The issue is not which method, the issue is, what should we apply to best address a question in its most stark terms with the greatest reliability and validity, and can we do it well.

Everything I'll talk to you about in terms of the details of some of the studies I'll now bring you in a minute, are studies that ask the bottom three questions. What is going on? What do children need to learn to read? What is it? Descriptive kinds of studies, epidemiological studies, correlational studies, just designed well. And let's say we've matured in terms of the what, and we want to ask, can we change the what? That is, does something effectively move behavior, can we have an impact on learning in the classroom? Is something more effective than something else? Then we can no longer rely on descriptive methodologies in total. We can model some things, using some descriptive methodology, but we've got to begin to do X and Y kinds of things. Does X cause Y? And if we design a study that looks at whether X causes Y,

whether it's a teaching approach or a pharmaceutical, the job of the design is to make sure that everything between X and Y, that could also cause Y, is accounted for. Bacon's rule, straightforward, very hard to do.

Now, when we're doing the effectiveness question, as we do at NIH in a number of ways, and now at ERI, what we are confronted with is a number of possible designs. We can use randomized designs, which are quite strong, and definitely all those things between X and Y that could also cause Y are now washed out across everybody you're working with, by randomizing. But, we're stuck with a lot of less gold standard designs, quasi-designs where we have to control for things experimentally, and statistically, I mean.

Be that it may, the job is still to try to apply the most robust methodology to the type of question. In the studies I'll be talking to you about, we had to be able to describe the situation, we've had to see whether factor X worked more than factor Y in changing a child's learning, a child's behavior. But, we also have to understand why that is. We also have to look closely at process. What is it, for example, let me be practical -- let's say you're asking a question in downtown D.C. with kindergartners of whether or not certain instructional approaches are more beneficial in promoting reading behavior than others. Given the characteristics of the kid. Let's just say that's the straightforward question.

The theory that drives particular types of interventions or instructions might say that for kids at risk in Washington, D.C., given a foundation of language ability that's a bit at risk, we are going to have to teach using direct, explicit, systematic kinds of approaches. Another theory might say, that's not really the case, those kids don't need direct, systematic approaches, they don't need to be taught letters and sounds. Why? Because that way, in fact, demotivates them. What they need is information presented a more culturally relevant way, bigger picture kinds of things, and so forth.

So we can adjudicate that. We can get after that. And we can take research into Washington, D.C., and actually address this question. For which children are which instructional components and combinations thereof most effective in which phases of development. And we can find that we can address that question, but we're going to leave quite a few kids behind. No one approach is equally effective. And I'll take you through that in a minute.

So how do we do it? How do we do all of this? How do we answer those four questions? And this is only one way to answer four complex questions, and it clearly is not the only way, but it's an expensive way. I will say that. In order to figure out what it takes to learn to read, why some kids have difficulty, how you prevent reading failure, and how you can remediate reading failure you have to be able to study large numbers of kids, and that is what we have done for many, many, many years. And as we study these children, let me try to computize everything I'm saying now.

When we're doing the what kinds of studies, the descriptive kinds of studies, we have to sample children from all walks of life, from every SES, every ethnicity, every race, we have to bring them into the various studies at about five years of age, and that's still too late. But all the

work I'll be talking to you about now that reflects the principles of science that we've gone over so far, are longitudinal, prospective studies. And the reason they are that way is because when you're studying children's learning, you have to be able to start studying them before they get into the learning situation. How else are we going to do horse and cart analyses? Now, there are ways to do it, but not as informative as what I'm talking about.

So we have to measure thousands and thousands and thousands of young five year olds across all hypotheses, or hypothesized characteristics that may go into reading. We have to measure language, perception, memory, attention, we have to measure home life, we have to measure books in the home, we have to measure how much parents read to kids in bed, or on the lap. We have to measure all of that, all of the contextual factors, all of the direct factors. Then we have to follow those children for at least five years, average nine years. And quite a few of these 5 year olds are now in their 20s. Measured three times a year at 44 sites, 9 sites in Europe and Asia, all on the same core battery, although that changes as we refine, and so forth.

So what happens with the what when you can do that? The what says that now we understand that if we measure something at 5 years of age, and we watch reading behavior develop at 6, 7, 8, 9, 20, we can have an idea of what it was that was so critical to reading that may help us identify kids at 5 and 6 at risk for reading failure.

You couldn't do that unless you put in place longitudinal designs that measured the array of domains that we think or we hypothesize theoretically that went into learning how to read. Now, given that we can do that, and given that this population of kids that we're looking at, 39,000 kids, a lot of those kids become extraordinarily good readers, many of them don't. And so one can look at the end of the first grade, and second grade, and third grade, and see with some degree of confidence what it is that predicts reading behavior at the end of those epochs. So now we can predict which kids are at risk, and we can do the second level of analysis we talked about, and that is we can bring to bear experimental designs, or quasi-experimental designs.

We can use the information from the what studies, we can identify kids at risk in Washington, and Boston, in Seattle, and Toronto, and Tallahassee, and Houston, and so on and so forth, who are at risk for reading failure, we can randomly assign those kids to different instructional approaches, and we can, indeed, start to understand what it is that teachers might be able to do at a very concrete level with five, and six, and seven, and eight year olds that mitigates against failure, that actually produces strong response in these kids. We are far enough along to do that.

And to cut to the chase, if you ask what does it take to learn to read, we know that it takes a good deal of background information that the kids bring to school. But, if they don't have that it places a premium on very strong teaching, because most kids at risk in this country, of that 38 to 40 percent that don't read, come from poverty. And the reason they're at risk in learning how to read is because they didn't have the types of interactions, the quality types of interactions linguistically, or with literacy materials, that gave them the foundation. So here they come into preschool and kindergarten and so forth, bereft of these foundation capabilities. And if teachers do not know how children learn to read, then it is going to be difficult for them, as hard as they're

trying, to be able to pick up the slack. It's going to be hard for those teachers to identify what it is about the kid's entry skills that makes it difficult.

The problem is, we know what it is, to a large extent. We're going to learn more. A lot of what we know will be revised, no doubt about it. But we know that learning how to read requires a series of abilities that have to be in place, albeit, they are not sufficient in their own right. So we know that learning how to read requires phonemics awareness, the jargony term of the century. Everybody has heard phonemics awareness, right? It simply means the ability to understand that the words that come by one's ear are composed of smaller units of sounds called phonemes. What makes it difficult for kids is the ear never hears the teeny sounds, you ears don't, mine don't, because when we talk with one another we don't spell out oral language.

What makes it easier for middle and upper middle class kids to come into school and learn quite well how to read no matter what the approach is that they've been read to a lot, they've played with language a lot, they've sung nursery rhymes, they've read Doctor Seuss, and all of those are wonderful foundational capacities, or capabilities to help kids understand that language actually can be played with, manipulated and segmented. But, kids who don't have that are substantially at risk, and so what they're going to require when they get to kindergarten and first grade, as we found, is a clarity in the instruction that indicates to them that the language actually has teeny pieces to it, and why are those teeny pieces important, because those teeny pieces have to be applied to the letters, and letter patterns. The phonics piece, I hate to use the F word, but I'm safer using it with you guys.

[TAPE CHANGE.]

MR. LIONS: There's no doubt about it, you cannot learn to read an English alphabetic language, or any alphabetic language without the sound structure and the ability to apply sound to print. You can't do it. Some kids need more clarity than others in learning it, that's for sure. But, if you just teach, for example, phonemics awareness and phonics to our nation's children particularly those at risk, we're going to screw them up, because it's necessary, but not sufficient. And this is where we get into this phonics, whole language nonsense. Where is the press table? Get off of it guys. I mean, really. If you're ever interviewed about this you can take them through, and I'm sorry, I'm paining the press -- you can take the press through a wonderful mosaic, show them all the data in the world that reading is a complex process, it requires phonemics awareness, it requires phonics, it requires fluency and the rapid application of those skills to print, it requires vocabulary, how are you going to understand what you're reading without having a vocabulary, it requires reading comprehension strategies, it requires all of that. And they'll come back and say, isn't that a phonics approach, or a whole language approach.

Again, the continued use of phonics-whole language, qualitative-quantitative dichotomies serves as a proxy for low intelligence in the asker. It's dumb. I mean, look at piano playing, do you want whole piano playing or piece piano playing? How about basketball? Reading is a skill. It's an accumulation, an amalgamation of other complex skills that must come together in an

integrated fashion so that, for example, if we're doing studies down in D.C., and we're trying to figure out which instructional approach works, and we find that, for example, one is hitting 60 percent of the kids, but it's leaving 40 percent behind, how is the teacher going to figure out how to modify the thing that works for the other 60 percent for the other kids if they can't ask themselves, how does a kid learn to read?

Phonemics awareness, phonics, fluency, vocabulary, reading comprehension strategies, a lot of motivation. So, guess what, they're going to have to bolster the program to make it stronger along those dimensions that the kid is not getting. That's research to practice. Now, am I talking out my whatever? No, because we know that we can train teachers to do just that. And I'll tell you, one of the most, not serendipitous, but clear findings we had from the D.C. study was that you can compare programs, but indeed what makes those programs most effective is the teacher preparation that comes along with it, is the coaching in how to use the programs in the classroom.

These kinds of comparisons, or trials we're doing, got us into trouble with the business community, though, because the business community wants to take hold of the data and say, my god, we are showing good, scientific results. In fact, when we tried to do clinical trials using published programs, they invariably, not all the time, but invariably or most of the time will turn into evaluation studies, because guess what happens. Even though we randomly assign, there are spies out there from the business community, they know when one program is bumping a little bit more somehow, and they will come in and revise the program in the middle of the trial, good business, lousy science. And, in fact, in D.C. we found that one approach that if we had just let it alone probably wouldn't have been effective, but the teachers were trained to ask that question, what does it take to learn to read, and they took an approach that wouldn't have been as effective and bolstered it, and they made it more clear and more direct, and that approach did quite well.

So it's not either or. But, there are there clear opportunities to bring, obviously, what we give to teachers from a knowledge perspective, and what these programs are doing. So what does it take to learn to read? It takes all of those kinds of items, but it has to bring to bear to make the translation into classrooms vibrant and real. It takes an enormous amount of preparation with teachers about what exactly we're talking about. And I think we simply all of these things when we talk about reading and education.

What's coming up quickly are just trials showing you that, in fact, different approaches have differential effects with different kinds of kids. These data are available in a nice form. But, we also have to figure out, as I tried to explain in this last set of examples, that we not only have to describe the kids, we not only have to assign them to do effectiveness work with different kinds of designs, but we have to figure out how the instruction is implemented, the process of that, the ecological contextual work that makes youngsters get the picture.

Now, all of the instructional studies we do combine descriptive qualitative designs, randomized and quasi designs, as well as ethnographic qualitative designs at the same time. They're very expensive. The average cost of these studies is about \$1.5 million per year direct, if that means anything to folks, very expensive. And they go on for at least five years.

We also bring to bear another series of disciplines and questions that tries to understand what is it about the body, the physiology that may make learning different in different kinds of kids. And in a number of smaller studies within this network we know understand how best to image brain as it's working in children, without invasive kinds of properties. And we can look at kid's physiological signatures in those systems that under good language and reading, before, during, and after different types of intervention. We've got 11 sites working on this, and the data are converging. What is good, just to digress a bit, is when you have these many sites one tenet of science about converging evidence is met. Reliability is met, if we're getting similar results across the sites, and convergence is met, in terms of amount of data.

If you look at the right hand side you're looking at the left hemisphere is a seven year old, seriously at risk for reading failure, and not reading well at all, about the 14th percentile. This is a kid from Houston in Jack Fletcher's shop. And this is a youngster that's going to go into trials in Houston. And what that right top, left hemisphere shows you is a bubble of orange in there, the brain doesn't do that, as you know, it's just a way to code, in this case, magnetic activity that happens when blood moves in certain ways. But, what it shows is what we call a deactivated signature for knowledge and reading. Those systems where you see a little bit of orange should be highly activated in a strong reader. It's a very reliable signature in right-handers.

This is not a genetically driven learning disability you're looking at. This is where the world or the environment came to this youngster without a great deal of language and literacy interaction. And thus did not exercise those systems in the brain that need to pull sounds out of words. Didn't need it, kids can talk okay, you don't need that to speak. But, the kids can't read because of that, or associated with that. And lo and behold, after 70 hours of intervention, in this case luckily with something that banged up nicely against the kid's characteristics, we've got now a reader at the 61st percentile, whose brain is displayed down below, who now shows a normalized activation signature for those basic foundational skills in reading. Behavior change rides right along with physiologic change. And that helped us test the idea that certain systems in brain, or hypothesize that certain systems in the brain were, in fact, instrumental in unpacking linguistic streams. And the convergence on this finding is substantial.

But be that as it may, the practical kind of thing that you're going to see in legislation, that you're going to see driven through the new peer review systems, or the modified peer review systems in the government is that as people as asking for federal money, what the reviewers will clearly ask is, have you asked a question that's important and significant, that can be falsified? Have you selected a design and a methodology that, in fact, meets the question's intent? Have you been able to ensure that the measurements you're using to look at all of these child properties are, in fact, measuring what they think they're measuring and reliably so. And have you attached to all of this a linkage between a theory or hypothesis, or methodology, or interpretive mechanisms, and given that you've done that, are you now ready to submit it to external critique. And therein lies science.

It doesn't have a great deal to do with methodology. It has to do with the quality of the methodology, and the appropriateness of that methodology for the questions asked. It does have

a lot to do with people knowing what they're doing. And I started out this talk talking about how practitioners can come from any, or many colleges and universities that prepared teachers with entirely different subsets or sets of knowledge. And, indeed, in the educational community, that is the case with respect to research training. Where our educational shops many times are polarizing and training people along the either/or dimensions of qualitative-quantitative, when in fact, the type of person and teams of people that will ultimately solve the problems that we have in front of us are those that ask just those basic questions that I've taken you through, and hopefully somewhat clearly.

Thank you.

(Applause.)

MR. FLEISCHMAN: Reid, I think since you were a bit shy and noncommittal on where you fall on this stuff, I was wondering whether you might take one or two questions.

MR. REID: Absolutely. Yes, sir.

QUESTION: We have a program we think works. We don't have a lot of money to do research. We certainly can't hire a third party. So how does that get done? I mean, I know how it gets done. The question I really want to know is, is it...?

MR. REID: You know, this goes to a couple of levels. Number one, so you have a product that may be helpful to schools, programs and so forth. You're going to have to demonstrate to peer reviewers that that, in fact, has been found to be effective. This is why it's important to remember that the review is going to be peer review. There are no lists of programs. There are very few programs that carry gold standards with them.

One of the things that the peer reviewers will try to make sure of is whether or not the program that you're recommending actually has been based upon, developed upon what we know, if it's reading, about reading. Does it cover phonemic awareness, phonics, vocabulary, fluency, and comprehension? Does it do it in a way to motivate kids? It ties, obviously, to those definitions in the law.

So, if you don't have gold standard randomized trial evidence, the first line of thinking is, does it sit on top of what we know about reading at this point?

The second line would be, have you accumulated any data showing progress of students that have interacted with the program? The third would be external trials. The fourth would be randomized trials. And I'm giving you may take on these kinds of things.

The second question you ask is, how does one who is developing programs and so forth, develop the capacity, the internal infrastructure to do this type of thing? Clearly, you'll see the legislation moving toward helping people, publishers and developers create in some ways at least the knowledge that they have to begin to do their own research and development, and they have

to allow that internal research, just like pharmaceuticals, to come under the aegis of external review. That is going to be an investment, but obviously business drives this. And those programs that have the ability to have or develop this infrastructure to look at their products in development, making sure they're honing it, and then rendering the final product to certain trials, and having that feedback to greater improvement is going to be critical in the next decade. We should expect nothing different from publishers than we do pharmaceutical companies.

Now, the pharmaceutical companies, we provide a good deal of capacity, and some financial kinds of recourse, and some regulatory mechanisms to help them save money, and we can start to look at those same ideas with providers, with vendors and so forth. But the issue here is, when you have a body of knowledge, you have gone through the criteria that we just have talked about, and Lisa told me about this morning, and we know that if we're talking about reading that a program should encompass these things, then it behooves us to make sure that that product or program has been tested, because the external validity side of this is critical. If a program is developed, not yours, I'm not sure what yours is, but not making this a point, if a program is developed in Sheboygan, Wisconsin, and then applied in Washington, D.C., and doesn't work, it may be because the populations were entirely different. So, when schools call us and say, should I adopt this program or that program, we can't endorse any programs. We can say which ones have gone through trials, which ones have particular evidence. What we can say is, make sure that that which you are buying actually applies to the kids you're teaching.

It is going to be tough. How do we begin to change a culture? We're going to have to try to support growth development in the best ways we can. We're calling in publishers daily to talk with them about these things. Smaller business folks shouldn't be penalized because they're smaller business folks. But, I mean, it will be peer review that drives this selection of programs and teacher preparation interactions more than anything else. And we are asking for, as you can see in the legislation, a good deal of creativity on preparing teachers. You know, if you look at what the characteristics of teachers are, I just want to get this across, that leads to changes in outcomes for students. The top one is teacher's cognitive ability. The second is domain specific knowledge about what they're teaching. The third is general knowledge. The fourth is experience. The fifth is master's degrees. The sixth is in-service training.

So when we're starting to figure out where to spend money, as Lisa Keegan was talking about, it's no longer in the process of what people know and how they know it, it's in the outcome that that process leads to. That which should be purchased is that which leads to benefits on behalf of kids.

MR. FLEISCHMAN: Reid, thank you very much.

(Applause.)

MR. FLEISCHMAN: We're just debating who goes first. Don't go anywhere yet. We just have a few more comments to make on this incredible day.

And, Reid Lyon, thank you so much again for bringing a very provocative and informative

presentation. Reid Lyon.

(Applause.)

MR. KOHLMOOS: Before you go, I would just like to make a few very short summary comments about today. The first comment I have to make is to say thank you so very much to all of you for taking the time out of your day to come here for what I think was an auspicious event.

(Applause.)

MR. KOHLMOOS: For those of you who got here late, we talked about the auspiciousness of this occasion early on this morning.

And second of all, I would like to thank three people, three people who put together this effort in a very practical way: Sara Mead from Progressive Policy Institute; Ann Heald from the Education Quality Institute, and my very own John Waters from NEKIA. I would like to thank them very much for all their hard work.

(Applause.)

MR. KOHLMOOS: As far as a summary statement, I have two comments to make. One is, I think it's time to take the question mark off of the title. That now is the time for education research to, indeed, be the leading edge of school improvement.

The second summary statement is that, whereas research should be on the leading edge, there's a whole continuum of activity that occurs from the time the research is done until it reaches the classroom, and that's something that we didn't really talk sufficiently about today, and that we haven't built in our conversations today a full enough connection. And I know some of you are wondering out there about what that connection is, and I would say that's for the next forum.

Thank you.

MR. ROTHERHAM: Sure, thanks, Jim. Thanks, Steve.

Again, I want to echo and thank our host, our gracious host there, the AED, for the use of this terrific facility, and thank NEKIA and EQI for their partnerships, and second what Jim said about Ann and John Waters and the work they did, and especially our own Sara Mead. And I actually just want to take a minute to also thank Renee Rybach (sp), Julie Radocchia, and Kathleen Porter from my team for their tireless work at PPI on this project as well. It took a lot of all three organizations and a tremendous amount of sweat equity to pull off what we did today.

The three concluding comments I have, the first is this issue that was raised by a couple of people, Mike Cohen raised it, David Myers raised it, and so forth, in terms of the sophistication of practitioners as consumers. That is clearly a paramount issue, and it should be a takeaway from

this event to start thinking about what can we do in terms of preparation, particularly in the schools of education and so forth, to make sure that the high quality work that we're talking about, that the actual users are sophisticated consumers, and the work that Steve is trying to do in that regard is terrific. I think that's one thing.

The second thing I think is a takeaway saying that the dialogue and needs to continue. As I look around the room, there's a lot of representatives of business and the private sector here. Not only people involved actually in the education sector, but business more large, and we need to have a discussion about how to involve different stakeholders in this dialogue and this debate as it moves forward, and we should be very careful not to sort of isolate this issue.

And then my final thought, as Reid pointed out here, and people pointed out throughout the morning, your data are nonpartisan. They're not Republican, they're not Democratic, they don't have particular ideological leanings. We too frequently assign them to them, but data are nonpartisan. They're frequently contradictory and confusing, but at the same time there are some things we know. There's clearly too much politics in this debate right now. I think a number of our speakers raised that, Dave Myers raise it, Checker raised it during one of his questions, Sharon talked about it with regard to this whole debate over teacher quality, Lisa talked about it with more of a light touch, but there's clearly too much politicalization here, and if we can take one thing out of this room today, it ought to be to try to move this debate forward with a little more of a focus on what's important here, and a little less of a focus on the politics, which clearly has an absolutely corrupting influence on the process.

MR. FLEISCHMAN: I also want to thank everybody that was involved in putting this together, and just quickly move to one point, which is, if we're going to take the question mark off, and if we really are going to make sure that research becomes the leading edge of effective school reform, we're also going to have to, while we do this, remain purposeful but modest about what we expect to occur. This is a very long-term effort, and we're, I think, fairly early on in this. I think that I sense a commitment from all kinds of folks to support this, support the notion. I think that's one of the things Andy was just addressing. And others have heard me say this before, but in doing some of the research we've done in trying to understand how you turn research into practice, I came across a citation about the practice of hand washing, which as the CDC reports is the single most effective way to reduce the spread of infections. And the New England Journal of Medicine in 1992 reported that after training individuals working in an emergency room, and then monitoring their hand washing behavior, they never got a compliance rate of greater than 50 percent. That points to the difficulty even when we know what works, and we conclusively know that hand washing reduces infections, the translation of that into practice for whatever reasons is very difficult.

So that's an argument for being purposeful, modest, and sticking to it over a long period of time to get the changes we need.

MR. KOHLMOOS: Thank you so very much folks for coming.

(Applause and end of event.)