STEM EDUCATION RESEARCH DESIGNS:
Conceptual and Practical Considerations for Planning Experimental Studies

Workshop Theme

FRAMING THE KEY ISSUES

Friday, October 12, 2007

Westin Arlington Gateway
Arlington, VA

A Claims-Based Approach to Evaluation

Anthony E. Kelly, Professor
George Mason University

Introduced by Michael E. Martinez, Professor
University of California, Irvine

MR. MARTINEZ: Okay. Let's reconvene. We have stayed remarkably on schedule and we're going to keep up that pace. So our third speaker of the day is Anthony E. Kelly, better known as Eamonn Kelly. He is a graduate of a somewhat questionable program in educational psychology at Stanford University.

[Laughter.]

MR. MARTINEZ: To which I admit I'm a member of the same program. He has also spent a number of years at Rutgers University, and as he shows here on his slide, he's intermittently been a program officer at National

Science Foundation. It looks like there is still a connection. There's still a tether to the NSF. Professor at George Mason University.

We're very pleased to have Eamonn be our wrap-up speaker today, and just so that you know, he has prepared two versions of his talk, kind of feeling his way through the day, a Dr. Jekyll version and a Mr. Hyde, and I've been pushing him toward the Mr. Hyde version. So there we go. Professor Eamonn Kelly.

MR. KELLY: Let me thank Mike for the invitation to come here today. I don't know if this is a reflection of how long it takes after funding for something to get done, but when this award was made, Michael was an associate professor at UCI and how he's a full professor at UCI. Got promoted last week so congratulations, Mike.

Yes. This is the more radical of the two. The more conventional one would have been to take the Educational Researcher article that you got as background reading and form it into maybe a checklist or a set of guidelines for doing more broad evaluations of programs than the standard one.

That paper is written ostensibly as an attempt to strengthen structured abstracts, but the rhetoric underneath it is actually about evidence and warrants and claims and causality.

So, as Mike said, there were two versions of this, and last night over drinks, he suggested it was a good idea to go more radical because you were

going to be coming in after lunch and you deserved some entertainment for your resilience.

So what I've got is this one. It's a Toulminesque perspective, which you'll have seen in that Educational Researcher paper, but with eschatological overtones. In deference to Larry Suter's sensitive nature, there's no scatological overtones.

MR. SUTER: That's what I was referring to.

[slide 2] MR. KELLY: Okay. You can relax. In this paper, Kelly and Yin, 2007, and in the paper that I wrote for the Journal of the Learning Sciences, there's an attempt to move away from method towards argumentative claims, and the reason for that was that unlike the hard sciences or how the hard sciences are represented, for example, the analogy is often used in medicine that randomized trials work in medicine, there's a lot, there's very little hermeneutics when you take an antibiotic; it just does what it does, and that's fine.

But as you're beginning to work in socially constructed environments, it's not as clear how, what exactly represents the medicine, what's the problem and what's the cure. The causal structure isn't perhaps as isomorphic as one might imagine.

So instead of going back into thinking about responding to the strengths and challenges facing randomized trials, which there's plenty--there's some very good books out, Kirk among many others--this was an attempt to

reconfigure the debate in terms of rhetoric and claims rather than terms of method.

In Toulmin's model, knowledge is about making claims, not necessarily about answering questions, but making claims, and those claims are based on some data, and the data are nothing without--a machine that works--grounds. So a particular piece, a particular datum doesn't mean anything unless there's a larger grounds from which that gains some of its meaning and perspective according to Toulmin.

And then behind grounds, there are larger commitments to backings. This, Toulmin has written extensively on this, and I won't go back over it, but what has become more pertinent in educational research circles is this question of warrant. And Dewey talked a lot about warrant, and in a lot of educational research findings including qualitative findings, you hear a lot of this influenced that or this caused this, but the warrant for that claim isn't always very clear.

So this is the argumentative structure that Toulmin talks about and has developed it quite extensively, and in education, educational research, as in many areas of science, the warrant is tied to a research method and that research method is part of a larger research methodology, and the "ology" being the logos or the meaning and appropriateness of using the model that the method instantiates.

So the question that was raised in both those papers was what is the proper backing for method or does method get us where we want to go in terms of claims?

So I've decided here not to reprise the Educational Researcher paper but to extend a little bit into what is evidence, the notion of evidence, and think about that in terms of a larger view of research methodology in something like education.

[slide 3] Here's some of the eschatological references. This is the Great Chain of Being. On the bottom you've got--this is from the National Research Council, 2002, Shavelson and Towne. On the bottom, what is or what is happening? Then association, what else is happening--a co-descriptive claim? Then if you apply numbers to things, are there unusual numerical patterns? Then is there something systematic happening? And you can look for what appear to be causal mechanisms. Cause is a slippery construct.

Joe Maxwell wrote a nice paper in I think 2004, I think it was, in the Educational Researcher on different views on what causality is, but there is a purported mechanism, there's a time dependence, contiguous action, things like that.

And then up at the top, I'm going to call it mechanical causality, but there it is. Either by structural, laboratory, physical or experimental controls, and experimental controls including successful random assignment and a successful experiment, what Maxwell calls variance causality, does it appear that X causes Y?

And the thing that differentiates the top from the second from the top is this notion of a counterfactual, which is if X causes Y, would Y have occurred if there wasn't an X?

So if you take X out, so that essentially is your logic for placebo and control trials. So you want something where the medicine is not being given.

The logic of randomized trials makes an awful lot of sense. It goes back to Fisher who was criticizing in a number of literature reviews a lot of the studies that had been done on the effects of fertilizer, and somebody supposedly came to him and said can you just suggest something other than criticizing the studies because he kept claiming that there was a selection effect, and it dawned on him that if you randomly assigned, then you would get protection from alternative hypotheses in terms of the claim that you could make.

The idea makes sense. It's been around for a long time. It's shuffling cards before you play poker. Right. You want to get rid of a selection bias. You want a fair chance to lose your money. Right. So that's the simple logic of it.

[slide 4] In the National Research Council report, there is this linkage between question and method. So it outlines those things you saw in the prior slide, what is happening, is there a systematic effect and so forth, and then it asks are the methods appropriate to answer the questions and to rule out competing answers, and notice that what we've got here is the linkage between methods and questions.

And methods and questions is fine, and I'm just thinking that, I'm going to suggest that the linkage may be a little too tight because if you believe you've got a causal question, then certain methods are privileged. If you think you've got a descriptive question, certain methods are privileged, and I think I want to get free from the automaticity of that linkage.

So, for example, if somebody says I want to look at the impact of some program, that impact suggests something to do with causality and then it kind of makes sense that you would if you could, and Jimmy showed us a nice example, do a randomized trial. Quasi-experimental designs, of course, is also possible.

Mike said he went to a very good workshop given by Cook and Shadish that has some very exciting new work being done in quasi-experimental design.

But noted in this formulation here is that the question is causal in the sense of X causes Y, and that X causes Y is actually a direct quotation from NRC 2002. That's how to describe it, X causes Y.

The question is how do you think about impact in something that is pertinent to schooling and to what extent is that question--under what conditions is that question reasonable?

[slide 5] So let's look at a model of stages of research that might be active when somebody is doing program evaluation or something as broad as MSP, which is $600 million and growing of interventions, and not only there, but in

other portfolios, both at IES and the Department of Education, NASA, NSF, so forth.

[slide 6] You can't read that from here. I can hardly make it out. But on top it says Evidence. Evidence is data in service of some claim. Evidence is data in service to a claim. All right. And that claim can be relative to the stage of research.

So let me tell you what's going on here. In the middle you can see this box right here. Is it systematic and so forth? So the NRC questions.

Interestingly, left out of the NRC question--I couldn't find it--what does it mean? The question of meaning is absent even though there's a lot of discussion of meaning earlier on in the book. When they get down to what the questions are, that one seems to drop out of the analysis.

But you can do observational studies, correlational studies, experimental studies, implementation studies, diffusion studies, and then the communicating of your results in many different ways to many different audiences. And by the way, these are not necessarily tied to any particular size of intervention. You can do these in very small studies.

So last night over drinks, I was explaining that I did a randomized trial on the proof of trigonometric identities with high schoolers from an Aberdonian academy--so you talk about the non-representatives of that sample--looking at the viability of states-based analysis as a way to help children understand trigonometric identities.

So you don't have to be doing the large program in order to do a randomized trial. An awful lot of stuff that used to be done in Ed psych was randomized trials around particular questions.

So all of these various methods that are available to somebody in educational research are not necessarily in competition one with the other, and you can apply most any of them at any level of analysis from micro to macro.

So this is the typical set, and under Observational, by the way, you can't see it here, but this says mixed methods in each case. So I'm not necessarily restricting things to quantitative or qualitative analysis. I think we're at a point now where each informs the other.

Okay. But other things are happening. For example, this is what is happening, and that's what the NRC report is talking about. Well, what could be happening? Right. Or as Bob says, you know, what could work better? Not what works, but what could work better?

So what would happen in this thing if we intervened? What is this thing? This thing is a school. Well, that's work that is going to create an innovation. An innovation by definition is something new. It's going to perturb the system. So you do some brainstorming and development; you create a prototype; you iterate the prototype in order to make it better.

And while this cycle continues, this part of it here is the area that's getting a lot of attention under design-based research methods, design research, and others. In fact, if you want to see a more developed model of

this, Brenda Bannan-Ritland who's sitting in the back right-hand corner has a paper in the Educational Researcher 2003 which makes this model clear and sophisticated.

But the area, where does a randomized trial fit in here where you're trying to engineer something in an iterative fashion, and you are working towards an end in view, but you're not quite clear what the variables are, you're not quite clear what the measures are, and your sense of success is in terms of a dialectic between the teachers and the students and yourself as a researcher and the design team.

And a lot of the rhetoric here or the claims are claims around innovation and design processes, divergent thinking, creativity, risk-taking, that kind of thing.

Then, as you move on, if you've got something that looked very good, then you would do some sort of confirmatory testing to look at its local impact.

And oftentimes from here to here would be where you would see typical program evaluation, and even though there hasn't been much money for it, IERI did give money to resource large interventions, kind of looking at the effect of a program that was successful locally at scale.

And the methods for doing this are similar to but different from here. The costs are different. The representativeness of the findings is also an

issue in terms of how you select from a larger population in order to make this claim.

But I might suggest that traditionally a lot of, say, curriculum development work was having some idea of a reasonably formed innovation and then looking at it and doing a formative and summative research and maybe in some cases applying it more completely.

It happens around the world that things all the time are going to scale or not going to scale on their own, and the things going to scale or not going to scale are factors related to adoption, adaptation and rejection of innovations, and as Rogers points out, this is a social process. Right. Diffusion of innovations is a social process. It's a process of people agreeing to so it's not causal.

It's not clear in what sense it is causal, but it's people agreeing within a practice to take on a new practice, and Rogers has identified a number of variables in terms of, for example, the complexity of the innovation, the opportunity costs for giving up your practice to take in this new practice, trialability--can you take a sample of it--observability--can you see it in action?

And he has got, there's a lot in the diffusion of innovation literature talking about S curves of adoption where it takes a long time, many, many years in order to get enough people to do it, that you reach an inflection point. People who are opinion leaders take it, and then you get a lot of acceptance or the S curve dies.

I believe the time period for the adoption of kindergarten in the U.S. was 50 years it took it to become properly accepted approach in schools, and now practically everybody running for governor of a state is suggesting a kindergarten or pre-kindergarten education, but it took a long time for that to occur.

And when you think about those time scales, the idea of judging whether or not an intervention has been successful is something that is hard to know for many, many years.

Also, in the diffusion of innovation literatures is the problem of what's called the established base. Right. So if you are in a practice, you're a teacher, you're a faculty member, you're working at the National Science Foundation, you have a way of doing things, and that's the established base, which tends to work against innovation, and one of the strongest established base problems in education is the textbooks, the textbook companies.

There's a very small number of them. And often so much of the educational activities are based around textbooks, exams at the end of the textbooks, slides based on the textbooks, and so forth. And I believe that many of the state exams that have been written have been written by the testing arms of textbook companies using, you know, IRT.

So there's a lot of activity working against, for example, science inquiry where science inquiry is more open-ended, and it may be the case--looking up here, long-term policy consequences--if you actually get an innovation in place like No Child Left Behind, what are the long-term policy

implications of--Jimmy said--I don't know if you were exaggerating--75 tests; is that what you said? Was that an exaggeration? I hope it was an exaggeration.

MR. KIM: Reading First.

MR. KELLY: 75 tests. It would be interesting to see the effect of a treatment of removing the 75 tests and letting them read instead, but that's neither here nor there.

So anyway so you've got policy studies that go on, thinking about the long-term impact because as you know from complexity theory, the behavior of a system should be looked at over the long haul rather than just looking at an impact. So you might get an immediate impact which is positive or negative, but you got to let the system settle down to decide what's truly going on.

And depending upon the impact on this, and it's going to interact with science education now if No Child Left Behind goes into science education. I was at a meeting with Congressman Baird, who is a psychologist, who was complaining about the likelihood that science is going to get reduced to memorization because it's too hard on paper and pencil to authentically represent what science is.

And if that happens, then all of this armamentarium that you see here is going to be in service of what? Right. So when we complete the circle here, you come back to is there an interpretation or reinterpretation of the central constructs, again, back to the larger claims for the society? And, of

course, all of this here suggests that we need ongoing design assessment, construct validity work because it's not a given. Temperature is a given. Weight is a given.

Meaning is not a given. So in this area here, overarching you've got hermeneutic studies of everything that's going on there.

[slide 7] Claims and evidence. Education is a social activity. We didn't always have schooling. Schooling was a human construction for certain goals, and the goals vary, and based on how the goals vary, you will have indicators that do and don't support those goals, and if you are doing an innovation or an intervention, you may or may not be in align with these goals.

For example, STEM goals, there's a difference between achievement and learning and learning to learn and skill and cognitive apprenticeship.

Achievement, which is a standard one, is a snapshot. It's a one-time point. You know SAT. In this state properly called the SOL test. There's a one-time point. You know, summative. Did you make it; did you not? But of course learning is really, if nothing else, change over time.

So another STEM goal would be to look at change over time rather than single-time point. If you do this, then the measures are more complex and the research approach has to be more complex.

The learning to learn goal, this is the transfer of learning idea that's been around since Thorndyke, and a lot of what we hope is that children

will learn something in school that they will take with them somewhere else and be successful and that they will meet all the challenges that are coming down the pike.

That's a hard thing to work for and you hear it all the time, and the Journal of the Learning Sciences had a nice issue that Joanne Lobato was an editor on that looked at many of the current models of transfer of learning including Bransford and Schwartz, Preparation for Future Learning, in which they complain about what they call sequestered learning and testing.

So we expect people to work in teams socially, but we test them alone. They're not allowed to cheat. They only have a small number of resources, and this somehow, this behavior is somehow a good selection from the pool of behaviors that you're interested in, in the workplace or whatever the case may be.

Well, depending upon where you come down as a school or a state board of education on this, you're going to do different things in terms of the activity the students get involved in and the measures that you're going to use.

We can't ignore enculturation, political core values like social justice, retaining identity and tradition. That's certainly one that's important in an immigrant community like the U.S. I'm from Ireland, lost most of it. But the Fiddler on the Roof notion, that there's something important about tradition. This is causing upheaval in France at the moment, and in England for that matter, and in Holland.

So other goals could be repair. If there's something wrong with the school, we want to fix it. The school is doing okay. We would like to incrementally improve. We want to adapt and adopt innovative practices. We want to sustain gains. We want to foster resilience and that might be actually quite a good goal in some schools.

How about excellence? You hear a lot about it. Singapore is an interesting example because it has convergent-attained excellence but is looking for divergent excellence, and by that I mean I had the honor of going over there in the summer of 2004 because they were topping out on TIMSS, as Larry can tell you, topping out on TIMSS which they see as a convergent indicator of learning. So they were teaching essentially to the test, and getting, they were topping out on the international comparisons.

For the engineers, the people who mostly run that country, that wasn't good enough. Being the best in the world wasn't good enough on that type of measure because with the explosion of economic development in China, Singapore has a rounding error, and you need to do something other than demonstrate the ability to do well on tests.

And what they wanted was they want to make Singapore to be the bioinformatics capital of Asia, and they built, the DNA lab for Nanyang Technological Institute was built attached to the School of Education, and all the teachers are required to go through and do DNA samples and learn all about genomics.

And they were interested in this kind of divergent thinking--how do you take a population of students and make them leaders in something like bioinformatics? It's an interesting question. It's not very clear what the goal is, what the variables are, what the measures are? How do you come at it? And they were interested in design research, this kind of innovative iterative approach.

And what they ended up doing was they ended up having the scientists make little DNA toolkits that were given to teachers and were sent to all the schools, and then the educational researchers were brought out to watch the dialectic unfold between the little piece of technology and the students and teachers to see what it was they needed to do to reengineer this toolkit to be successful.

So economic problems--sorry--economic problems. I don't know if you read the most recent The World is Flat? But we're living in a global world and anybody can upload to it. You don't need many resources. So, the economic goals the school might have had in the '50s, you know, when Detroit was flying high and going to fly high for decades, and the goals they would have economically now might be very different, and based on the economic goals, you're going to see differences in approach.

Then you've got, of course, the so-called extracurriculars. I call them extracurriculars because I've got two kids going through the Virginia schools, and as the demand for testing goes up, art drops out, music drops out, sports drop out, and then you've got this emerging phenomenon, extracurricular

phenomenon that I see with my kids of they live in a Web 2.0 edutainment kind of a world.

My daughter has no idea what a library is. She uses Wikipedia, and what impact this kind of extracurricular--it's what Robert Yin might call a societal rival that's going on--how that intersects with it is an interesting thing.

In any event, there's an argument that we've got a complex, socially constructed, historical, political, cultural, networked, but distributed, globalized and personal world of schooling, in which the claims may exceed X, whatever that is, causes Y, whatever that, and the concept of evidence is controversial.

I'm going to suggest that claims emanate from belief systems that define the interpretive context in which data can become evidence. Right. Data, you don't find it lying around. You don't find evidence lying around. You need frameworks into which that becomes evidence or not, and we'll get more to the eschatological in just a second.

[slide 8] So here's an example of a poor school. I mean a school that has a lot of demands on it. Any idea why the little red schoolhouse is red historically? Iron ore, iron ore based paint was the cheapest paint so red was--they didn't have much money for schooling right in the very beginning.

Okay. So the school has to worry about the individual student. Now that individual student, you could say an awful lot about, but they're

coming from some environmental factors, physical, emotional, health, nutrition--there's a lot of factors that you're assuming--also for groups of students.

And in some sense, the goal is the short and long-term memory of the students. So you teach them things; they remember them. Here's a simple model and they wander off. Of course, that's based on the brain which we're learning more about. That breaks up into involuntary and voluntary processes. Some of them are subliminal but trainable, like biofeedback, for example, and some of them are completely unconscious like healing or antibiotic action within your system.

And then you've got the voluntary where we put most of our--this is awareness, cognitive processes, self-regulatory process, in order that the person will learn, but this over here is still going on.

So, for example, you might have a child who's under a lot of stress. So here's a whole lot of these variables that you have to consider, the school has to consider, and then on the other side, you have got the fact that once you get a social system or people, you get groupings, and when you get groupings, you get emergent and regulatory behavior.

So political activity. Right. So you've got Jefferson on one side and Martin Luther King redefining over long period of time what "we, the people" means, and I could have put Abraham Lincoln in there.

You've got economies and incentives that has an impact on what teachers do and parents do and where kids see themselves going. I've got a--off the record--got a son who I thought I was going to be in music or drama who is now at Virginia Tech doing business. Triple majoring in business, as he says, "cha-ching."

So you got globalization and then you've got the human as a toolmaker. This is the AK-47. It's the most recognized gun in the world and is kind of symbolic of revolution. Primitive tools. Computers are tools, hammers are tools and so forth, and these kind of link up now in the Internet in interesting ways.

You have got military service, religions. Anybody know who this is? It's Yusuf, used to be known as Cat Stevens--"Morning Has Broken." Right. So I put him in here for the music value.

Dance, languages. We live in other regulatory systems as well. I'm putting this down here in the bottom because it represents not only communism but history. Recently I was in a bookstore and one of the students from the local high school came up and said, you know, I've seen this on books; what is it?

Isn't that interesting? I mean they grew up in a world that is post-Iron Curtain. In any event, this is all going on and all important, as is this, as we're thinking about where we're going to go in terms of evidence and claims and causality. Exactly where does causality fit in this?

[slide 9] So let's take an example here of paleontology which is nice STEM content area. So if somebody is interested in teaching that. So what is evidence in the social and historical context? And let me take this example. So here's the claim. And here is method. And this is why I want to kind of separate method which is a kind of, it's important--it's an important series of steps related to the structure of argument that allows you to say with some protection that the claims that you are making are less assailable than might otherwise be the case. Right.

[slide 10] So they've recovered the sample. They've identified the age of the rock that it came from. They got physical and digital records. By visual inspection, they said this is an example of the fossil of a fern.

[slide 11] And then going up the chain of being, the associative claim: these fossils are associated with rocks that are millions of years old. We've got a new method of radiocarbon dating which confirms the age so we've got a correlation. We count the number of fossil ferns. We correlate it with the age of the rocks.

And then we claim without a counterfactual that these ferns co-occurred with sediment that formed during the rocks, "proving" their age and their fossil nature. Right. And of course, fossils only appear in sedimentary rock.

[slide 12] And this all looks fine and good. How about the gold standard? Well, hard to do. You have to go back in time and randomly assign, difficult to do. If you did, actually, according to this guy--he was an Irishman, James

Ussher, you go back as far as the 23rd of October--we're about to celebrate the founding of the world happened the 23rd of October 2004 B.C. You know the part of the Bible that says "somebody begat somebody begat somebody," well, this guy did it, and added on some other data and decided that this was the year that everything started.

[slide 18] He was also a bigot, but we can leave that aside. However, you look at this thing here, and this also looks frost. Sorry, this also looks like ferns. All right. So you're looking at that saying "hmmm," I wonder what's going on here. So somebody has said that those things were ferns, and you don't like the idea of fossils in the first place.

[slide 14] So you say, well, maybe from the paper this might be a test analytic rival or what you might call a construct validity concern.

Is the description, descriptive assertion, this is a fossilized fern, is it really a fern? So there's a rival claim that it's not, and the specific rival is that it's due to something else. It's due to concretion, and concretion is a process in which minerals are extruded in rock.

[slide 15] So you go back and take a look at it, and it turns out if you reanalyze this from a chemical point of view, it turns out it's not a fern. This is just like frost. So as the chemical process is depositing, it turns out there you essentially got Brownian motion here, and as things adhere, they're more likely to adhere to something that already has something adhered to it so you get this kind of fern-like structure.

[slide 16-18] Okay. So the people who are interested come back and say, okay, well, I don't think this is concretion. I think this is really a plant. So here's something else that looks like a fern. So this is a counter-assertion, and this doesn't look like it's due to concretion, nor does this, and this is Dinosaur National Monument, and here would be big bones, and you can see the size of this girl here.

So it looks like there is something going on that isn't--this does not look like a concretion problem.

[slide 19] And when you put all that evidence together, you get this biostratigraphy field which looks at layers like this and sets them up in terms of time, and then it looks at the fossil record and decides this I guess is closer to the real chain of being, and of course people are way up here somewhere.

This is evidence for people, some people, that what you have got in the claim is that, you know, the planet is very old and we evolved, and that would seem to be noncontroversial except the controversies are real and they continue so this--don't you remember this episode in 1996? There was a Mars meteorite that people thought that this was an organism, with an electron tunneling microscope--I don't know if you remember that.

[slide 20] Bill Clinton went on the news and he said today this rock speaks to us across all those billions of years and millions of miles. It speaks of the possibility of life. If this discovery is confirmed, it will surely be one of the most stunning insights into our universe that science has ever uncovered.

And then even as it promises to answer some of our oldest questions, it poses still others even more fundamental.

Right. So here we are teaching paleontology and you get this kind of response. Here's a rival. Methods prove nothing. Ontology and epistemology trump methodology. All right. And this biostratigraphy that you're talking about is really contrary to the true world view. In this case, I'm representing the Young Earth Creationists who are Ussherites. They think that everything was under 6,000 years.

[slide 21] So what do you do with these fossils? You accept that they are fossils, but this is what you conclude: that God planted the fossils to test your faith. Right. So it's an interesting thing about evidence. You've got these things that people are agreeing are fossils, but what are they evidence of?

The evidence goes back to the larger framing in the claim. It doesn't go back to method and it doesn't go back to that which you can see clearly in front of you.

[slide 22] Okay. This gets more extensive. There was a guy in 1857 called Gosse, and he had the Omphalos hypothesis, which is the Greek for navel. Adam was born with a navel, and from this he concluded that because the world works, because it's functional, everything had to be created at the same time; otherwise it couldn't work.

And the fact that it's here and it's working shows that there is no evidence that we can ever get of the age of the earth and the universe. You can't talk about the age of it. It all happened together.

This was extended into the argument that we were all born five minutes ago with all of our memories intact, which is true, of course. Why I put you guys in my memory bank, I don't know, but, okay, but what you see here, now we've gone from the claim that started off as scientific to absolutely unfalsifiable. How do you falsify the claim that we were all born five minutes ago with our memories intact, which is the extension of this argument here?

[slide 23] Right. So again, just, this is the eschatological. I'm just taking the point just to draw out the notion of evidence and warrant and claim maybe to an extreme, but maybe not too much of an extreme. So, in summary, learning and teaching occur in complex social, historical and cultural context. Evidence is defined within and animates the larger semantic and semiotic context in which people live. That says schooling is happening where people live.

What to teach is often controversial. In some cases like phonics and algebra, you see algebra all the time; you see phonics all the time. This complexity is reduced and randomized field trials with narrow standardized tests can indeed be persuasive.

However, if you look at some of the reading comprehension literature, it's not clear that the measures don't fall out as easily. You get interpretation, you get hermeneutics, you get a lot of problems. Algebra, thanks be to God, is still just algebra.

So, in general, teaching STEM content generally can be controversial. We've seen a geoscience example. There's a woman called Victoria Richardson who is at the University of Michigan. She was in as part of the cyberinfrastructure discussions, and she said that she had met some geoscience instructors from high schools and colleges in Texas and there was a big push not to have geoscience in the curriculum because it went against this larger view held by some vocal people that the planet wasn't that old; yet, you had to train the students to know enough science to take the oil out of the ground, which apparently was put there because God liked Texans and I guess he liked Arabs, too.

So you've got a real controversy in this case here. The E is for engineering. There's some very nice articles coming out now about the engineering in New Orleans. I mean the whole approach that the Corps of Engineers has taken led to the problem. You didn't need to have the disaster that you had. So engineering sounds like it might be something that's not tied to values, but it very quickly tied to values.

In fact, apparently some of the canals they built for economic reasons acted like the mouth of a funnel to take all that water that was coming in and shoot it right up into the Ninth Ward. Also, the destruction of the wetlands that were sitting out in front of New Orleans was an economic and political decision that had a big impact. So engineering even can't escape it.

The Internet is under--I'm going to put it under technology. There's a lot of positive effects of Internet. I've got kids. There are some

lousy effects of the Internet. The Internet is also related to the spread of democracy and maybe other not so attractive things. But again it gets controversial quite quickly.

And many schools have restrictions on people downloading things, and now that cell phones are getting Web-enabled, that's becoming even more problematic.

In mathematics, I was looking for one and I found one, nonlinear dynamical systems. Why is that controversial? Well, a lot of that modeling is used in global warming, and global warming is also controversial, and let me finish with this final eschatological example.

[slide 24] If you look at belief systems and evidence and nonlinear dynamical systems, this was written in 1814 by Laplace: "We may regard the present state of the universe"--this was after Newton but before chaos theory--"We may regard the present state of the universe as the effect of its past and the cause of its future." An intellect"--was called Laplace's demon--"which at any given moment knew all of the forces that animate nature and the mutual positions of the beings that compose it, if this intellect were vast enough to submit the data to analyses could condense into a single formula the movement of the greatest bodies of the universe and that of the lightest atom. For such an intellect, nothing could be uncertain, and the future, just like the past, would be present before its eyes."

This world view within science was dead by the end of the 1800s. But this view is a deterministic view suggesting, because of Newton, we could just roll time forward and roll times backwards.

This is 2006. This is Pope Benedict. "Whereas the physical cosmos can have its own spatial-temporal development, only humanity, strictly speaking, has a history, the history of its freedom. Freedom, like reason, is a precious part of God's image within us and it can never be reduced to deterministic analysis. Its transcendence vis-a-vis the material world must be acknowledged and respected since it is a sign of our human dignity.

"Denying that transcendence in the name of the supposed absolute ability of the scientific method to predict the condition of the human world would involve the loss of what is human in man and, by failing to recognize his uniqueness and transcendence, could dangerously open the door to his exploitation."

Back again to the values. Interestingly, the premise that Benedict has here in 2006 is exact same premise that Laplace has in 1814. I guess he's not keeping up on the chaos theory work.

[Applause.]

[slide 25] MR. KELLY: Any questions, queries, theories?

MR. MARTINEZ: Plenty of time for questions.

MR. DELLA-PIANA: Sir.

MR. KELLY: Gabe, yes.

MR. DELLA-PIANA: If what to teach is controversial, which obviously it is, how do you move from your intervention to a comparison or an argument for why that intervention? So if I'm teaching history, do I want archivists to teach and make records? Scholars to make descriptions and predictions? Activists to learn how to change things? Storytellers? How do you move from your intervention, first to justify it, to begin with, not for its effect, but conceptually, and if you believe in comparison groups, then how you would choose them?

MR. KELLY: I think the problem is, is that if you're choosing a controversial area of history, it's going to be socially negotiated what's acceptable, and it's going to be negotiated above the level of the researcher is my guess.

I was at a talk at the National Academy of Sciences in '95, I think it was, and there was a guy from England over there, and he was talking about a thorny problem like the one you're describing, which he described as, quote-unquote, "the Irish problem." He was talking about Northern Ireland.

I spoke up and said in Irish history, that was always described as "the English problem."

[Laughter.]

MR. KELLY: So I mean you've got a notion about values and commitments and history and tradition that defines the content. Right. And

then I mean it's a peculiar thing. Is it an Irish problem or an English problem? I don't think a randomized trial with outcomes is going to decide that for you. I mean it's not a question that's subject to method; right?

It's something that's due to hermeneutics and negotiation and probably wars at some stage. I mean there's a sense in which the kind of mechanical intervention that--I have nothing against randomized trials--I was raised on the things, but the places in which the model works because behind every research method is a model, and you've got a model that's sitting behind randomized trials.

The isomorphism between the assumptions and the randomized trial and the phenomenon it's describing, the tighter that is, the more credible the claim. All right. So if I've got something--if I have plants and randomly assigning plants has no real impact on the plants. There are no social groups among the plants, and if the plants are essentially independent one of the other, I can talk about their average state before the intervention.

I give them fertilizer. That's up to me and then I decide what's growth, and then I can do a comparison at the end and get a T-test, for example. As the isomorphism between that example from Fisher, coming to the ones in medicine where the model is kind of falling down, as you saw with some of the medicines that are now shown to actually be killing people--Vioxx and others--and you've got problems with people who are not in the sample taking it. So there's a lot of cases in which you don't have the same control that Fisher had.

And then as you move out more into the socially constructed realm, then the isomorphism falls down in many ways. What you decide the construct is is controversial. What you decide the measure is is controversial. What you decide the intervention is--right--so I mean there comes a point at which you step off more into a hermeneutics space than into a mechanical space; it's more organic.

And I think, I can't imagine coming up with a control group for the Irish problem that would be at all satisfying to an Irish person regardless of how many zeroes followed the decimal point and the P value.

MR. MARTINEZ: Yes, Elizabeth.

MS. VANDERPUTTEN: So what kind of study could you do in science, for example, that could take into account both your very large conceptualization and the need to show, well, I guess, could you do any impact studies given your theory?

MR. KELLY: Yeah, I think so, but I think there's going to be a prior agreement on what's ruled in and what's ruled out, and there's going to have to be some negotiation between people's view of what the desideratum is, what the goal is, and the ability to, quote-unquote, "measure it;" right.

So what you've got in Singapore is the launching of a social experiment. You know these are children in high school. Will they be successful? Well, they're going to wait 15 years to see if these kids actually

started companies or did they all emigrate to Silicon Valley and never come back?

All right. So you have a notion. I mean in this what's studiable, I think an awful lot of things are studiable. A lot of it isn't controversial. I'm pushing on the controversial in order to pull the debate away from method and back into evidence and claim.

But there's a lot of stuff that you can do in terms of, for example, the study that I did on trigonometric identities. So in the pantheon of work, there's lots of observational correlation experimental implementation, communication studies that can be done that are going to be quite reliable and defensible.

And then there's a growth in this work here, and that we don't have--this methodology hasn't been worked out as well as it might be, and the responsive assessment to something as dynamic as this or the responsive assessment to something like this hasn't really been worked out and it's work that needs to be done.

So I mean I'm sorry I can't give a clear-cut answer because I don't think--if you can give a clear-cut answer, you have probably managed to reduce an awful lot of the complexity in the problem, and--by the way, I just want to say that I want to not suggest a randomized trial. I don't want to suggest randomized trials are inappropriate or the causality is not appropriate to schooling. That's why we send kids to school.

We think it's better. I mean the counterfactual is they're running around wild so I mean the fact that you put them into schooling is a treatment. It's a decision that this is better than the counterfactual of letting them run around.

So we've already decided that there's a causal mechanism, and I think that as we align to what is going on in schools, we can certainly come in at any one of these levels and help the schools get more efficient.

I guess what it points up to is Toulmin's comment about the difference between pharmaceutical research and medical practice. So pharmaceutical research is essentially chemistry. In public health, in the actual practice, he calls it a clinical science, meaning it's enacted with real people and real places with all of the caveats that come with that. So I think the parallel for education research is probably closer to public health in a clinical science than it is to pharmaceuticals and mechanical science.

So I mean I think educational research and schooling is more on that socially constructed side of the continuum rather than on the hard science side of it.

So the assumptions, the methods and the models that you pick up, the porting and the isomorphism, has to be considered as you move across the continuum.

MR. MARTINEZ: Eamonn, I'd like to jump in here too because I think that statement along with I think Larry Suter's earlier question about

where is all this going, what is the dividend here, I think there's a really big question there. And it also pertains to what this graph is all about, you know? What is the ontology here?

And it might be that you and I have different opinions about this, but on the one side, the dividend might be, okay, here's a bunch--here are different experiences that seem to work. They're effective in producing outcomes we care about, and I think, well, I think many, if not most, of the projects supported in EHR probably fall into this camp.

The result is: Here is an approach that works, and it happens to be our approach. It's proprietary. We've been working on it for a long time, and we've got good effect sizes and that's great.

Contrast that with a more knowledge building, a more theory building over time, and my feeling is that that is underplayed and the ontology of that graph ought to shift more toward not crystal palaces, as Soupees said in the 1970s, but more stable enduring truths that link what we know about students, student experience, and their development, learning, and success later in life, you know, more theoretical claims, and I'm just asking for your comments on that.

MR. KELLY: Just thinking about some of the words you used in terms of success and output, I mean there's effect sizes. If you look at the top there, the design research part, if, for example, I take some particular intervention for improving the understanding of algebra, and I've got--I

compare it against business as usual, and I get P less than .05 and I've got an effect size, that's great. That is really great.

The design research question is from what design space was that intervention taken? All right. So what were the assumptions within the design space that this particular thing was taken; right? So, for example, a lot of interventions in algebra up until probably Kaput got involved were books, you know.

There were worksheets and I mean--it was essential papyrus was the technology of choice, and it wasn't until Jim Kaput got involved with looking at dynamic representations, the most recent one MSP is now supporting, GeoGebra, which splits the screen and you've got algebra on one side and geometry on the other, and that allows you to teach the two pieces of STEM content at the same time; right.

Well, the fact that you can teach them at the same time makes you wonder where is the design choice justified to teach one after the other? Why should they be taught serially? And if they're taught serially, which one should be taught first? Should you teach algebra first or the geometry first?

Should you teach them together? Right? So even after an extremely successful intervention, right, there are still the design questions left open as to how expansive was the thinking that went into the thinking about making something better and different?

And I think you wrote a paper on the effect of the format of the test itself on the timing--you have a patent and figure responding, I believe; right? So they said there are ways in which the way we think about our ideas are already constrained by the design choices other people have made that predate us.

I was an evaluator on Heinz-Otto Peitgen's Florida Atlantic University project, and he as a mathematician was very angry at the Bourbaki group. There were a group of mathematicians in France who decided that they were going to make mathematics as pure as possible, to reduce it to the fewest number of symbols, and this was a break from the mathematical tradition which had been highly ornate and with lots of diagrams and stories.

They wanted to take all that and squeeze out, which they did quite successfully. And that coincided with the printing press and the cost of printing mathematics from the old school or the new school, it was much more efficient to do it under the Bourbaki model; right. So these very sparse symbols were taking over the more complex ways that mathematicians used to communicate mathematics to one another, right, which begat the textbooks because the mathematicians' books became the inspirations for somebody to write something that was kind of for a high school audience.

So, you know, I'm working with my daughter on trying to uncover the meaning packed into one or two symbols, and it's just--it's a cognitive demand that understanding now what we know about cognitive science, we wouldn't do this.

Right. But the fact that it's there, it makes it almost invisible. From Rogers' point of view, it's part of the established base. And then it's hard to get schools to abandon that because that is what we do. So Kaput has had huge problems in trying to--well, unfortunately now deceased--but the SimCalc Project is doing its best to--you know, it's got some money from IERI, it's going to look for more money to do--have the big impact on schools.

But the notion of that being a durable effect comes back against parents' understanding of what mathematics is and, as Peitgen points out, a lot of that is historical accident that has to do with technologies and their limitations, not necessarily mathematics in its pure sense. And he, of course, waivers in the direction of geometry and fractals and chaos and beautiful pictures, and he thinks that's a far more powerful way to teach.

In fact, he would claim that you start off with fractal dimension and move from there to Euclidian dimension, not the other way around, and that if you did that, science would be much easier for kids because very little of what we see is actually Euclidian. I mean this room is, but walk outside into the world, there's not much that's really Euclidian.

MR. BORUCH: To return for a moment to your declaration that some of the STEM contents may be more controversial potentially than others and including, and especially perhaps, for example, engineering and the New Orleans thing. In talking about failure analysis in the engineering sector or in the legal sector, values of the courts, evidence in the courts, and so on and so forth, and failures in the education sector, school reforms that, quote, "don't

work," no appreciable effect, I've been talking and trying to think about this topic for awhile including with colleagues from Singapore.

John Yip, who you may know, he's former high up in the Ministry of Education, and his take on it was to avoid the word "failure" entirely, rather to ask the question, quote, "What is wrong with being wrong?"

The implication being that we have to admit the possibility routinely of being wrong. Understanding how to handle, package, treat being wrong is then important. Understanding how, for example, in this so-called design experiment arena that is part and parcel of the activity, understanding how to take the onus out of, quote, "being wrong" means that some of this controversy can be removed is kind of important, potentially important.

And certainly this gets beyond--well, it gets again at this interface between the science, among the science, the culture and the politics. Some of you may know that the Chinese Academy of Sciences is--president of the Chinese Academy of Sciences is actively lobbying, agitating for a Chinese statute which tells Chinese scientists, quote, "hard scientists,"--those other guys, not our hard science, but their hard science--that it is not a mortal sin to fail. This is on account of at least ten or 15 of these guys being sort of brought to the "wall of shame" because it's announced that they falsified or fabricated data in the interest of increased prestige.

So, part of it, it's how do you arrange your own thinking about, quote, "failure" or being wrong in this context? You need to have some standard or benchmark for being wrong. It could be a counterfactual; it could

be something else. But how do drug companies handle this? How we handle it in education research? How the courts handle it? And so on. These things do vary. I'm sort of curious about the way you view it.

MR. KELLY: Certainly, in the, I guess it's the epistemology of mathematics, the notion of being wrong is a very bad thing. Right. There's one answer and this is how you get it, and if you made a calculation error and you're wrong, that's a problem.

Compare that to Richard Lesh's model, Eliciting Problems, which are far more--you start off with the assumption that you can't be wrong; you can make a better or worse argument for your case. So I think this permeates--your question permeates throughout. It's a sense of what the subject is and what it means to be wrong, and how long you're going to tolerate being wrong, the consequences for being wrong.

I mean you talked about all the studies that are never published that are sitting in the drawer because they in some sense were wrong or were failures. The notion of what is going to be a failure is going to be something that is going to be awhile to decide because something that is an apparent failure may over time prove to be actually the rule and over time then gets accepted.

I believe that the paper that started FedEx got a C on it. The paper that did Google at Stanford didn't score very high. I'm sure that guy is a little miffed that he didn't give that a higher score.

[Laughter.]

MR. KELLY: So I mean the sense--

MR. BORUCH: Wharton School at the University of Pennsylvania does not make mistakes like that.

MR. KELLY: Right. Yeah, yeah, it's great, isn't it? Two and two is four. So the notion of failure goes back to Toulmin's notion of backing. Failure means a violation of some prior assumption, and in science, presumably as practiced, things aren't characterized as failures; they're characterized as refinements that are moving towards a more comprehensive understanding. So you can define--on your primitives, you define that out of part of the negotiation.

Certainly when we're talking about science inquiry for children, we want to include this idea where many experiments don't work. All right. The kids didn't know what they were doing. They mixed up, got the solutions wrong, or whatever the case may be. So failure is a notion that seems to be tied to, has a moral valence to it that we would certainly want to get away from. If people are being shamed or imprisoned because there's a moral valence, that we want to get away from.

In a society that's looking for more of something, so we're driving test scores up, right, so it's a model of acquisitiveness across the economy including learning. Then how you have toleration for somebody who doesn't seem to be making that progress but really may be the next Google, in which

case you should be pulling the kid aside and saying here's an F, but can I get stock?

Also, the notion of failure is very instructive, and one of my personal big regrets is that I had an opportunity to interview Everett Rogers, but unfortunately he passed away before that occurred, and he came to one of the design research meetings, and he was full of stories, and it's a lifelong regret. It will be one that we didn't get a chance to unearth them.

Some of you may have heard this before and I apologize, but it's the instructive side of failure. He talked about, a company that had made this new fertilizer and had done the Fisherian randomized trials and this was working. This fertilizer was working.

And they sold it to Iowa farmers. This was in the '40s and I think early '50s, and they waited for all the sales to come in. Larry is listening with great interest here, coming from Idaho. But they were waiting with great interest for the boost in sales because this thing was going to really improve the crop.

And they waited in vain. The farmers didn't want it. And this surprised them and they went back and looked at the results and the laboratory results were right, and P less was .05, and everything was hunky-dory. But the sales didn't go up.

So they finally sent somebody out who would have been an early ethnographer and said what's up? And the farmer said we'd never use that stuff

again; it burned the crop. It's disastrous. And they took it back; they looked at all the data. It didn't burn the crop. It was just fine. So they sent the ethnographer back out again and it turns out that the scientist hadn't considered that the farmers were men and they didn't read any directions. They just took the stuff, put it on the back of the tractor as normal and just put it full strength out on to the fields and it burned everything.

So they decided, okay, well, we'll fix that, and they brought it back in, and they prediluted it, and that caused all these engineering problems, getting bigger vats and all this kind of carry on. But in any event, they took care of that, sent it back out again, waited at the end of the year for great sales. No sales came.

Sent the guy back out again and the conclusion was that it sometimes worked, and they couldn't understand that because the science was the science and it worked, by God. But the farmers told him it sometimes worked.

Actually, sometimes it also burned the crop, and sometimes it did nothing at all. And this was very peculiar. They went back and looked. The solutions were exactly like the solutions should be. They had no idea what was going on. Ethnographer was sent back out again, and he wrote back a report, one line, “tractors don’t have speedometers”.

AUDIENCE PARTICIPANT: Tractors don't what?

MR. KELLY: They do not have speedometers. So they're speeding up and slowing down, speeding up and slowing down, and waiting for awhile and chatting and the whole thing is spilling and burning the crop, and then they take off again and nothing gets--okay. So this implementation study, right, was a whole bunch of failures, but it was crucial for the application of the science in the laboratory to actually drive up the crop yields.

So I mean there's another case in which if you were to look at that, ostensibly, that was a big failure, but over time they learned an awful lot that was then going to be an important set of variables to make further fertilizer or whatever the case may be, interventions, scientific interventions, successful.

So it's probably like art, Bob. You don't know. You know you got to wait a couple of years to decide that that which was dismissed, as with the word "impressionism," you know is now priceless. So it's hard to say.

Yes.

MS. STOEL: I think that's a wonderful story, and it really explains a lot of what's been kind of rumbling around in my head that I'm not fully articulate about. But I think the Rogers' stuff is so to the point about this randomized testing Because, if we're looking for a solution, we're going to look and grab something that promises to be our solution. So if we stop with randomized tests and they only show some significant change, we're going to be very disappointed.

So beyond, I mean it's sort of like No Child Left Behind testing. If the kids go from 18 percent to 40 percent, we say "whoopie." But if my kid were in that group, I'd say, oh, my God, what's going to happen to this child? It's unsatisfactory. It's sort of the wraparound to the testing and to the examination that doesn't seem to be part of the conversation.

It's the going back. It's, as you said, it's having people looking to see that there were no speedometers on the tractors. It's really trying to figure out in the way that you want to sell, you want to produce something that's going to make a significant difference. And fooling around with just getting the testing just right, but not taking it to the purpose of the testing, why are we doing this, what do we want to know?

Is Success for All really only getting the kids to 45 percent or is it getting them to 90 percent is, to me, the question? And who is promoting in this political funding arena, who is promoting the contextual analysis that really keeps the big question in mind?

MR. KELLY: Okay.

MS. STOEL: And that to me is just really out there.

MR. KELLY: Let me rise in defense of No Child Left Behind and testing, maybe surprisingly. If the baseline in which you're working is innumeracy and illiteracy, anything that moves that into numeracy and literacy is a good thing for society. And if the measures tend to be overly rigorous and

petrified, but you're still getting that movement from illiteracy and innumeracy to numeracy and literacy, I say great.

It's money well spent. The cost to a society of losing all those children, and by the way, their children's children's children, the cost is extremely high.

MS. STOEL: You think that 40 percent is--I'm not arguing with you about that. I mean that's fine. I agree with you on that conclusion. That's not my point.

MR. KELLY: Let me channel Russ Whitehurst, if I might. I heard him speak at AERA last year, and he made a very fine point, which was a small effect size, you should consider that and its accretive effect, its aggregative effect over many years. So that effect size might actually be something that as it builds upon itself over a number of years is actually quite significant. And there may be threshold effects.

So if I cannot read, and you give me an intervention that helps me read, that opens up the entire world to me. So I mean again it's coming back to where the measure is. I agree with you. I mean, essentially what this entire presentation has been about is what Cronbach called construct validity. That's what the whole thing is really about; right.

So I mean you do have--you do have a notion about what the desideratum is and how you're going to define it and define it for whom and how you're going to spend money and so forth.

Let me just say as a father of a gifted and talented kid, I think that it would be nice to see--part of the problem with No Child Left Behind is no child far enough ahead. You know, I mean if you look at the projections of the cost of us aging baby boomers on society, I really want more Googles and I want them now, and I want an awful lot of economic activity to pay for my Medicare, to put it bluntly; right.

So I mean it would be good--I mean it's important to leave no children behind. It's also very, very important to take the brightest children and to have them recreate our worlds in the most amazing ways, and part of that is going to be basic research like that, for example, gave us the Internet and then all the economic value that has occurred out of that.

So I think that as we think about, say, the National Science Foundation, the portfolio of--I see Greg Solomon in the back who is an advocate of basic research, so we may have--we may have educational research for which the payoff is not yet clear.

For those people who might be thinking of applying--I see Janice Earle sitting here--those of you that might be applying--there's a rewrite of the program announcements, and the REESE program announcement has frontier areas, and I haven't seen the DRK-12, but I think it's also going to have some kind of frontier areas of research.

But one of the frontier areas is the neural basis for the learning of mathematics. And when is that going to pay off? It's not very clear, but it's probably worth pumping some money into.

Another area is cyberinfrastructure, which is the marriage of the supercomputing network together with groups, collaboratories of scientists, together with teachers, together with students, to radically change what we think--back to the construct--what is science and how is science practiced?

Greg is going to be enthusiastically supporting basic research in cognitive science, particularly around causation. And also there is a call for new approaches to assessment to get at the issues that you're raising, and fleshing out and creating new research methodologies; right.

So you don't this, you don't have these color-coded things in opposition to one another. You've got some way of helping with an idea of a construct holding the whole thing together so that you are helping the situation.

One of the places where it would be nice to see work done in assessment is to find something other than item response theory. I mean it's good as far as it goes, but it doesn't go very far. So it would be interesting to see other methods that allow us to not just rank students, but to understand students so that we can, you know, augment their learning in other ways. So that we don't just tell people they're not good enough, which that's the real failure I guess.

MR. MARTINEZ: And maybe to measure qualities that we can't measure--

MR. KELLY: Yeah. That's right. That's right.

MR. MARTINEZ: --now. And that we know are important.

MR. KELLY: Absolutely right. I mean if you look at the SCANS, S-C-A-N-S, it's a survey that's done by business every couple of years. So you go to SCANS in Google, but if you look at the things they look for in a graduate from high school, there's not a single one of them that you would recognize from the curriculum in the high school.

They look for honesty, punctuality, teamwork, dedication, creativity, collegiality. These are things businesses are looking for. Stuff that--you know, I mean we're back now to the question of the goals, the social goals.

Although, interestingly, because of the American Competitiveness Council, there's a lot of interest in the Business Roundtable to put more testing into the schools. I know they're advocating, you know, common algebra tests across states and so forth, and you know that will have some value and some, you know.

I mean the answer, of course, is the APA AERA statements on testing. You want multiple indicators and discussions about what construct is and how you value it, and you know the limitations. I mean tests are tools, right, and every tool is good for some things and not good for others, but maybe you don't want to be utterly without it.

MR. MARTINEZ: Okay. That will close our third session. We have one more in 15 minutes. Are we okay? All right. We will reassemble at three o'clock for our final hour with our three panelists.

Be thinking of questions. You can ask those aloud. So we'll break now and come back at three.

Great. Thank you, Eamonn.

[Applause.]