

## High Stakes Testing-John Merrow Podcast #24 Testing analyst George Madaus on the science of testing

JOHN MERROW

I'M JOHN MERROW. THIS WEEK'S PODCAST DELVES INTO THE CONTENTIOUS ISSUE OF HIGH STAKES TESTING IN EDUCATION. STANDARDIZED TEST SCORES ARE THE ALMIGHTY BOTTOM LINE FOR ENTIRE SCHOOLS AND INDIVIDUAL STUDENTS. WHAT MAKES SUCH TESTS SO APPEALING?

GEORGE MADAUS

Education has become a political issue. And one of the things that legislators or governors can do is they can impose tests. And they don't have to worry about what goes on in classrooms, they don't have to get into the messy details. They get out ... and they get numbers out that are quantifiable. So, it's very attractive. And it's cheap.

JOHN MERROW

THAT'S GEORGE MADAUS WHO HAS ANALYZED THE TESTING INDUSTRY FOR MORE THAN 30 YEARS AND IS A SENIOR FELLOW WITH THE NATIONAL BOARD ON EDUCATIONAL TESTING AND PUBLIC POLICY. HE IS CURRENTLY CO-WRITING A BOOK TENTATIVELY TITLED 'THE PARADOX OF TESTING OUR CHILDREN.' HE BELIEVES THAT STANDARDIZED TESTING CAN PROVIDE USEFUL FEEDBACK, BUT WARNS EDUCATORS OF THE UNINTENDED NEGATIVE CONSEQUENCES. LIKE TED SIZER, HE DESCRIBES TESTING IN MEDICAL TERMS. CAN YOU IMAGINE HAVING SURGERY AS A RESULT OF A SINGLE TEST? GEORGE WOULD WANT A FEW DIFFERENT TESTS AND A SECOND OPINION AND I IMAGINE YOU WOULD, TOO.

HERE IS GEORGE MADAUS ON THE LIMITATIONS OF STANDARDIZED TESTING.

JOHN MERROW

The concept of sampling ... and I keep thinking back ... there's always a kid in class who would say, "Excuse me, is this going to be on the test?" That person intuitively understood sampling in some way. All right, help me with sampling.

GEORGE MADAUS

Well, you have this very large domain that you want to make inferences about, third grade arithmetic. The number of questions I could ask about third grade arithmetic are just infinite almost. So I need to sample from that domain in order to get an estimate of how a kid would have done had I been able to give the kid the whole ... every question you could conceive of. One of the main questions ... one of the main validity questions is, does the sample adequately

represent the domain? And so that's one of the very first validity you have to deal with, does this sample of fifty questions adequately represent the domain, a domain that might have different facets -- addition, multiplication, subtraction, division. If your test has fifty addition items and no multiplication, division or subtraction items, it doesn't adequately represent the domain, to be very simple about it.

JOHN MERROW                      What ... what makes a test valid or invalid?

GEORGE MADAUS                    The descriptions or the decisions or the inferences I make from the test about you or about a group or about a school or a state accurate? Are they accurate? Are they correct? To what degree are they correct?

JOHN MERROW                      If I'm studying diving, what's a valid test of diving?

GEORGE MADAUS                    Watch people dive. Put them through a series of different kinds of dives in which you have experts rate the mechanics and aesthetics of the dive and you don't give them a multiple choice test on diving. You ... that ...that's an area where you need to get actual performance. There are only two ways you can test people. You can ask them to supply you with an answer or a performance or a product, or the second way is to select an answer, and that's the multiple choice. The multiple choice didn't come in until the first part of the 20th Century, 1910, 1912, something like that, whereas the other kinds of test situations ... assessments, whatever you want to call them, have been with us forever. So I can ask you to give me something, whether it's answer, oral or written, whether it's a product ... uh ... a painting, a ... a ... a chair, a ... a chemistry experiment. I can ask you to do something like dive or repair an automobile and those are about the only ways you can test people.

JOHN MERROW                      We've been sort of playing with athletic images. Everybody's going to ... uh ... train for the ten events of the Decathlon, but then you'd need to sample, because you wouldn't have time to do all the things. Is that a ...

GEORGE MADAUS                    You can't give a multiple choice test or an essay test on a Decathlon. You make them perform and some of the situations ... one of the ... one of the principle validity questions when can you substitute an indirect measure for a direct measure?

JOHN MERROW                      Okay, a direct measure is you do the thing that ...

GEORGE MADAUS            Yeah, a good example of this is you ... on a paper and pencil reading test, kids read a passage and they bubble in an answer. The most direct way to find out if a kid can read is give the kid a book, ask him to read and then explain to you what they read. That's direct. The other way is one step removed from that. It's indirect. Now what you have to be able to show ... that the indirect correlates very highly with the direct, before you move to the direct ... the indirect.

JOHN MERROW            That makes it valid.

GEORGE MADAUS            That's right. Well, that's one validity. That's one indice of validity. Validity is multi-faceted and it's not an either/or and it's an ongoing process. A thing isn't valid for all populations for all times. It's valid for particular uses with particular people within particular time frames.

JOHN MERROW            You said reading. Okay, the test is reading. You give the kid a book and say read and ask him or her to explain it.

GEORGE MADAUS            Right.

JOHN MERROW            That's time-consuming and expensive.

GEORGE MADAUS            Exactly. And the history of testing clearly shows that the more kids you want to test or examine ... uh ... the more you are forced to go into efficient indirect, easy to score, administratively conveniently modes of testing. You ... you couldn't test every kid in the state by giving him a book and having him read. It financially wouldn't or administratively wouldn't be possible. Now teachers can do it. Teachers do it every day, of course and yet we don't factor in teacher judgments in any of these high stakes testing programs.

JOHN MERROW            You lost me on that. We don't factor in...

GEORGE MADAUS            Well I mean if you take most state testing programs, a kid gets a number on the state test and the decision or the description is made on the basis of that number. And if it's a ... if it's a multiple choice reading test, teachers might have other indices of direct reading. Kid brings in The Boston Globe and reads the sports page and that kid can read with comprehension. Might not have done well on the multiple-choice test for any number of reasons.

JOHN MERROW            I always hear people talking about whether a test is reliable. What does that mean? Can you...

GEORGE MADAUS           It basically means you get a consistent score. If I were to give you the test again, would you get roughly the same score? Or if I were to give you a parallel form of the test, would you get roughly the same score. It's the difference between reliability and validity is consistency versus accuracy. Now you can't have a valid test that isn't reliable, but you can have a reliable test that isn't valid. If you give a first grader a twelfth grade test, that is not a reliable ... that is a reliable indice of how the kid does on that test, but it's not a valid indice of what the kid has learned in school.

JOHN MERROW            Okay, okay. The ... uh ... the ... the thing that people talk about now in terms of measurement is this idea of a standard error of measurement.

GEORGE MADAUS           Uh huh.

JOHN MERROW            It sounds as if what they're saying is that every test ... every score has some kind of error built into it. Is that...

GEORGE MADAUS           Associated with it. Testing from my point of view, I always looked on testing as a technology. It fits any definition you want of a technology, it has an underlying, it has underlying algorithms, it has paper and pencil answer sheets in scoring things. But it's a fallible technology and like all technologies, there are places where it can break down. Now some of the error might be built into the test. For example, they can ... an item that is mis-keyed can get on the test. That's one source of error. The test can be too long. That might be another source of error. It can be a hot day and the air conditioning breaks down. That can be another source of error. For an individual kid ... uh ... the kid had a fight with the parents that morning or is sick, is coming down with the flu. That's another source of error.

All of these kinds of errors affect what the kid does on the test. Some of them are internal, some of them are environmental and some of them are in ... built into the test, accidentally, we hope ... we know. And all of that contributes to the person not ... the person's performance not adequately representing what the kid knows or is able to do. And we can es ... there are ways to estimate error.

JOHN MERROW            But I think back on the elections. Whenever there's a campaign, they'll release the poll and they'll say, Madaus is ahead of Merrow 46 percent, 42 percent. And then they'll say...

GEORGE MADAUS           And the error is...

JOHN MERROW                   And the error is....

GEORGE MADAUS               Sure. Yeah. And I mean ... again, I ... I like to think of it also in terms of medical tests. I asked my cardiologist one time, what's the error associated with my cholesterol. She said, about 20 percent. So that got me going about, what is the other in other medical tests -- blood pressure, prostate cancer screening, breast cancer screening, a whole series of things. And there's error. In medical tests, they know what those errors are and what they do there that we don't do in education, they then go out and get other measures and they put all this information together and then clinical judgment enters in and a treatment is either given or may not be ... may be counter-indicated. In education, we tend to take that score and it, act on it and not necessarily get other indices in these high stakes testing programs.

JOHN MERROW                   And that's not the right thing to do?

GEORGE MADAUS               No, I don't think it is, because any test ... if you ... if you come up with a number on any ... any one of these state tests ... let me pick a number. You get a 220. I know right away that that can be sixty seven times out of a hundred ... that can be a 215 or a 225.

JOHN MERROW                   There's a range...

GEORGE MADAUS               There's a range, sure. You're ... it isn't this precision that people think it is. There's a range in which your true score falls. A true score is if I could eliminate all this and I could test you over and over and over again and I could estimate what your true performance level is, okay? It's a construct. It's an imaginary thing.

JOHN MERROW                   But it's real.

GEORGE MADAUS               It's...yeah, as real as constructs get. So when you get an observed score, then I can estimate what the error is, same as I can do with your blood pressure.

JOHN MERROW                   So I come home ... hey dad, I got a 220 that really could be lower by some or higher by some.

GEORGE MADAUS               It could be a false positive or false negative. 220 is the cut score.

JOHN MERROW                   But the state says, 220 is the cut score.

GEORGE MADAUS Uh huh, that's right.

JOHN MERROW So if I got a 219, could that as easily have been a 221?

GEORGE MADAUS Oh, sure. And ... and a 221 could as easily have been a ... a ... a 218. So that's a false positive, false negative. So sure, absolutely. One of the things that people don't see is the arcane underpinning of all this. There's something called the three parameter item response theory algorithms that they use to arrive at these scores. These involve assumptions, these involve rounding, these involve all kinds of things that it's like a black box to people. It's like people ... like myself. I don't know what goes on under the hood of my car. And but if it breaks down, I take it to a mechanic. In a test, we don't have ... it can break down but we don't ... we don't really know what ... how this thing ... how I arrived at that 219, all right? Maybe if I had rounded a different way or I had put in a different assumption into my program, I might have gotten a diff ... a slightly different result. Not a dramatically different result, but a slightly different result. And just to take a number and say, this is ... this is your score, this is ... it ... we know that it isn't the true score.

JOHN MERROW If we know, the people who make these policies know, that when they say okay, 220 is passing and 219 is failing, they know these scores are squishy.

GEORGE MADAUS Sure.

JOHN MERROW Why do they do it?

GEORGE MADAUS Well, again, you're ... you're into ... uh ... a series of issues. One of them is political. You ... you don't want to seem to be fiddling around with the cut scores. These ... these ... uh ... these cut scores somehow, as I said, get reified. That's one reason. The other reason is that you're going to get a lot of backlash when people really understand how this classification system works and they don't want to deal with that. And the other reason is that they again, leave out one of the most important informational things we have about these kids and that's teacher judgments.

JOHN MERROW They don't trust teachers?

GEORGE MADAUS No, they don't trust ... that's why we have a lot of these state testing programs. They simply don't trust teachers.

JOHN MERROW Okay. There's a long history of kids being pushed along...

GEORGE MADAUS Yes.

JOHN MERROW ...sold for promotion.

GEORGE MADAUS Yes, absolutely.

JOHN MERROW Maybe there's a good reason for not trusting teachers.

GEORGE MADAUS Oh, there certainly I think there should be testing programs. I'm not saying ... I mean, I think testing gives really valuable information on how schools and systems are doing ... particularly certain systems that traditionally serving under ... poorly serving certain populations of kids ... ESL kids, special ed kids, minority kids, poor kids ... uh, and tests information can throw a lot of light on that. Where the cut score decision ... one of the reasons it's used is ... is a stick that you can beat people with. But it's the kid you're making the decision on. We can get very adequate information on how a school is doing, or how a system is doing, or how a state is doing without saying, "If you don't pass this test, you're not going to get a diploma, or you're not going to go to the next grade." You don't have to do that. In my opinion. Obviously, this is partly philosophical. Uh, the arguments pro and con. But, I think that these tests can, uh, and should be used to judge and hold the system and schools accountable.

JOHN MERROW Hold the grownups' feet to the fire?

GEORGE MADAUS That's right, yes. You can get a measure, but there is error in it. And before you make a decision, you should do what a doctor would do. You get other measures of the same thing, put them together, and then come up with a clinical judgment about the kid rather than just take the number and decide that you're going to do something or not do something.

JOHN MERROW So, even if the test were accurate, you'd still say...trust the teachers?

GEORGE MADAUS What do you mean by accurate? Accurate has error. I mean, your ... your blood pressure test, or your prostate cancer test has error in it. So, I want other indices before I'm going to decide what to do. And I would want to do that in reading, too.

JOHN MERROW Now, you said it's fair ... it's a good idea to have these tests ... and to use them to hold schools and the adults in them accountable.

GEORGE MADAUS Uh-huh.

JOHN MERROW But if the tests have no meaning to the kids...

GEORGE MADAUS Well, that's ... see, that's the stick pot. A lot of people think that ... that ... if we don't put the pressure on the kid, then nothing is going to happen. And that's ... that's an argument with a moral dimension. Well, why should you punish the kid in order to ... in order to whip the system into line. And it isn't just pass/fail ... it's the difference between proficient and advanced, between needs improvement and proficient. Uh ... there's error all along that scale. And, um ... I think you should get the numbers. I think you can use the numbers, but I think you ... you need to use them with other part ... other kinds of information about the thing you're trying to, uh, make decisions about.

JOHN MERROW So, the question we're always trained to ask is quo bono ... who benefits? The way we're doing things now...who benefits?

GEORGE MADAUS Well, uh, politicians are ... test companies are ... uh, teachers in schools are ... and the teachers in schools with the standards based reform, they look at the standards and it helps them better understand what they should be teaching. It gives them good, uh, information about the level of ... of performance expected. All that's ... all that's to the good.

JOHN MERROW The argument here is that we need these tests ... A) they are ... enable us to have a meritocracy or identify the best and the brightest ... who do well on these tests and will get opportunities they might not otherwise have had.

GEORGE MADAUS (Laughs) In this country? We know ... we know who ... who isn't doing well. We don't need tests ... we don't need these tests to tell us who is having a hard time and who's in trouble in school. We know ... we basically know the answer to that. You can ask any classroom teacher and they can tell you to a fair thee well who the kids are that are having trouble in math, reading, you name the subject. We know that certain populations are poorly served ... that there aren't schools that aren't doing a good job for these ... we know that. And you're not going to test your way out of those problems.

JOHN MERROW What do you mean by that?

GEORGE MADAUS You don't have level playing fields. So, it isn't going to create a meritocracy. Uh ... and we're not going to ... we're not going to

solve the problem by pulling up the tree and looking at the roots every year and then planting it back again.

JOHN MERROW           The commissioner of education in Massachusetts, David Driscoll, defends the use of the high stakes MCAS test because, as he says and the law prescribes, kids have multiple opportunities. If the first time they don't make it ... they have four more chances.

GEORGE MADAUS       That's right.

JOHN MERROW       And the last two are targeted to give them even a better shot at it.

GEORGE MADAUS       They call not only for multiple opportunities, but for multiple measures of the same construct. Not just repeating the same test four or five times. Now, the problem with repeating the test four or five times, Commissioner Driscoll is absolutely right ... they do have these opportunities. We also know a lot of kids don't stay around, uh, to exercise those opportunities, they leave school. Uh, and ... that's a problem.

JOHN MERROW       Kids fail and get discouraged?

GEORGE MADAUS       Kids fail and get discouraged. And, uh ... so that's ... that's one downside. But even if they didn't get discouraged, it may be that for some kids, they can't demonstrate what it is you want them to demonstrate on that mode of testing. In another mode of testing, they might very well be able to ... show you or, uh, what it is you're looking for. And so we need to try to get other ... other indicators of what it is we, uh, are truly interested in. In addition to the multiple opportunities.

JOHN MERROW       So, multiple opportunities and multiple (Overlap)

GEORGE MADAUS       Multiple measures, right.

JOHN MERROW       Fair question. Multiple measures...like what?

GEORGE MADAUS       Well, I mean, uh, the MCAS isn't the only, uh, fourth grade math test around. I mean, there ... there are an awful lot of them around. Or, they ... you can go in ... for kids on the borderline, you can go in and get direct measures. You know, we're not ... we tend not to worry about the false/positives, and we tend not to worry about the kids in the proficient and advanced category. We're talking about kids, uh, in ... in the ... uh, around the fair/needs improvement category. And there, you might need to go in and say ... give a kid a book and say, "Please read for me." Or you might, again, start to

get teachers back into the process. We all talk about we got to improve the teacher corps. And ... we got to treat them as professionals. And we've got to get them back involved in the process of evaluating kids on these things that they have day-to-day information. They have a ton of information on kids.

JOHN MERROW

So, help me with this, though. Um, here's ... let's use a hypothetical student. Let's call him George Medaus. And the teacher says he does all right in reading and his grades have been good. And some other measure ... some other test. But, on the MCAST test ... he gets a ... 180.

GEORGE MADAUS

212.

JOHN MERROW

Or 212 or whatever. He's way down.

GEORGE MADAUS

Yeah, yeah.

JOHN MERROW

Should we say, "Well, we have these multiple measures, we pass George?"

GEORGE MADAUS

I think again, you put all this stuff together and you come up with a decision. There may be a very good reason he was 212. Again, he may be that kid that was coming in and, uh, was coming down with the flu and had a temperature the day he took that test. Uh ... we don't know that.

Education has become a political issue. And one of the things that legislators or governors can do is they can impose tests. And they don't have to worry about what goes on in classrooms, they don't have to get into the messy details. They get out ... and they get numbers out that are quantifiable. So, it's very attractive. And it's cheap. Relatively speaking.

JOHN MERROW

Lots of money being spent.

GEORGE MADAUS

I mean, given the overall education budget, testing is a very small part. It's getting bigger, but it's a relatively small part. It's not nearly as expensive as equalizing funding. Uh ... putting money ... money into ... in service training of teachers. A whole series of things that would cost a lot more money. And you can appear to be addressing the problem. And that's why I say you're not going to test your way out of the problem. It's ... if you're going to test, you're going to do other things, too.

JOHN MERROW The question has been raised about the accountability of these testing companies. There's not a great deal of oversight.

GEORGE MADAUS Right. Well, I've maintained for a long time, uh, that we need, um ... an independent body that does that sort of thing. There's plenty of places you can go to get a test built, administered, scored and reported. But there are few places where you can go to get an independent evaluation, either of the technical quality of the test or of the impact of the ... of the test. And the impact is, in a lot of ways, more important.

JOHN MERROW The tests themselves, George, um ... I wonder how ... how good they are. Uh, the other day a young student, a tenth grader, found a mistake on the math tests.

GEORGE MADAUS Uh-huh.

JOHN MERROW Someone else revealed that James Madison was identified as John Madison.

GEORGE MADAUS Yeah...

JOHN MERROW Is that an area of concern?

GEORGE MADAUS Oh, sure. That happens. I mean, there are these items ... that's what I mean about the limitations of the technology. These items ... these kind of items are going to slip through. And, um, it ... it can become, uh, a serious issue in some cases. We, uh, were involved in a, uh, some colleagues and myself were involved in a lawsuit in Alabama where they were certifying teachers. And we found an item called ... it was called a carrot item, and it was miskeyed. It simply was ... there were four choices, and the key was B and the right answer was D. And what we were able to do was go back then and find the kids who failed and didn't get certified by that one item ... and had D ... and we were able to actually identify those kids ... young adults ... and, uh, show how many there were ... I forget the number now. But, even if it's one or two, that's an important life ... and it doesn't do any good to say, "Well, they can take it again multiple times."

JOHN MERROW It's life changing.

GEORGE MADAUS Yeah, well, it ... it certainly affects you. Uh ... and some people are more tenacious than others, and ... will stay at it longer. But you run the risk of driving people out of the system. You need to look at

the populations that you want to help the most. And how do you help them the most without setting up situations where they might leave school ... or push them in the GED's or ... uh, things like that.

JOHN MERROW                    You make it sound as if these policies really do hurt a lot of kids.

GEORGE MADAUS                They can, yes. What I'm saying is that the technology that you're using to do these things has inherent limitations that we don't fully take into account. And we don't take into account other measures or teacher judgments. And if we could somehow put those together, then you might have a system that, uh, would be fine.

JOHN MERROW                    So, if George Madaus were in charge ... you'd use the test to hold the adults accountable?

GEORGE MADAUS                Yes. Yeah.

JOHN MERROW                    But you'd use multiple measures ... for the kids.

GEORGE MADAUS                I would use multiple measures and I would find a way to get teacher judgments back into the mix. And I would evaluate those teacher judgments so that ... so that ... there are ways that you can look at ... at teacher grading practices to see if there's inflation. Or ... or if they're, you know, giving automatic As and Bs ... you can monitor that. It might take you a little time, but there are ways that you can pick that up. And you could change that. You could go a long way to changing that system. Uh ... but right now, we don't do that it's not convenient.

JOHN MERROW                    Young Pete Peterson ... the tenth grader we found the irregular pyramid or ... um, it was the John Madison/James Madison error ... does that suggest that maybe there are other mistakes? Other bad questions?

GEORGE MADAUS                Yeah, I ... you ... there may be other ... maybe not as blatant as that, but there may be other ... what we call ambiguous items where, uh, A is the answer that they want, but B isn't that bad. And you could make a case for ... you might be able to make a case for B ... maybe a strong case or a weak case. But there are always ambiguous items, uh, that appear on these tests. And, uh, that ambiguity, uh, again, masks the true ability level of the kid. The kid may have read something into the ... remember, adults are the ones who write these items. Kids are the ones who answer them. My favorite example of that is the famous cactus question. Uh, the ... there was a ... third or fourth grade test. And they said: Which of the following needs the least amount of water? Cactus ... they had

a geranium ... these are in pots ... and then they had a cabbage. And of course, the adults wanted cactus. A number of kids picked the cabbage. And when asked why they said because the cabbage was picked, it doesn't need water any more. Perfectly (Laughs) perfectly sensible choice for those kids, but they got marked wrong. There's a lot we don't know about how young kids approach these items, what they read into these items. Uh, we ... it is ... this is not, uh, high altitude physics. This is a much more basic, rudimentary kind of technology.

JOHN MERROW

You're talking about the bad questions and ambiguity. But the technocrats would say, "Well, yes, but we're getting better."

GEORGE MADAUS

Uh-huh. Well, if you want an analogy, the testing technology, uh, to automobiles ... we ... in terms of the testing technology, we are still back in the Model-T era. Uh, now ... I think eight, ten years down the road when we find new ways to use, uh, the ... the computer technology and we meld the two together, and by that, I don't mean we just throw multiple choice items into a machine and have the kid do that ... I ... I mean, simulations. For example, uh, one of the exciting things that we're working on is, uh, with doctors and medical simulations. Now, pilots take flight simulators ... the military has ... so, I think that we're going to find more and more that K to 12 testing eventually is going to go down that road. But right now, the technology is a Model-T technology. And we do ... we're very good at that Model-T technology. But it's not ... it's not a Porche. (Laughs)

JOHN MERROW

The simulations that medical schools, the Army, other places use ... those are actually teaching...

GEORGE MADAUS

Sure.

JOHN MERROW

Whereas multiple choice ... the tests we give now are ... I don't know if they're gotcha, but they're not necessarily teaching.

GEORGE MADAUS

No, they're not. No. And the other thing is ... is the misinformation that the purpose of a test like the MCAS is for diagnosis. Uh, again, let's do a medical analogy. Uh, you come in and you take a test in May ... and I don't give you the results back until, uh, November. And then I don't even give them to the same doctor, I give them to another doctor. I mean, that's not diagnosis. If you want diagnostic information, you've got to give stuff with very fast turnaround and very fast feedback. And that's not what these tests do. These tests ... they classify people, that's what they do. Now, that's not diagnosis. Uh, you have a low score in reading. If I want

to ... what I really want to know is why do you have a low score in reading? That's what I really ... I mean, that's what a diagno... a diagnosis would be ... why are you, uh, scoring so low?

We know that teachers teach to the exam. You talk about frameworks, well what becomes important is the tradition of the test. You change the test in some European countries, the first thing the teacher says is ... they don't want to see the syllabus, they want to see the exam. Uh ... and so the standard becomes the test. And that way, you get teaching for the test. Now, people say if ... you can have tests worth teaching for. Well, no test is ... should replace a curriculum.

GEORGE MADAUS            These indicators are corruptible and tests are corruptible. And you can have these scores go up and not have the underlying learning that you seek to improve go up. Uh, so I mean, we ... there are wonderful examples of this going back a long, long time.

JOHN MERROW            This is an old dance.

GEORGE MADAUS            This is an old dance, right. And I predict that with what's going on now will start to ... implode.

JOHN MERROW            Implode?

GEORGE MADAUS            You can have a very good, um, goal mind, but the unintended consequences of those goals very often, uh, come back and ... and bite you. And that's one of the reasons we need to very carefully document impact ... document positive things like teachers using the frameworks and teachers having a better understanding of what ... what it is they're doing. And ... and those are very positive. But, we also need to look at the negative things. Are kids being forced out of school? Are teachers just narrowly cramming for the exam? We need to look at all of this and ... get a better understanding of, um, how these testing programs actually, uh, play out in classrooms with different socioeconomic levels, with different kinds of kids, and different settings, and so on. And ... uh, we tend not to ... do that. We're always going to have tests. And we're always going to have, uh, some kind of accountability ... we need accountability. But these high stakes things focused on kids, uh, I think that's going to start to change.

JOHN MERROW            THAT WAS GEORGE MADAUS OF BOSTON COLLEGE. I SPOKE TO GEORGE IN 2001, BEFORE PRESIDENT BUSH SIGNED THE NO CHILD LEFT BEHIND ACT, WHICH HAS BROUGHT THE EMPHASIS ON TESTING TO AN UNPRECEDENTED LEVEL.

HIS INSIGHTS INTO THE SCIENCE OF TESTING AND ITS INTRINSIC DEFECTS HAVE BECOME EVEN MORE PERTINANT AND STARTLING. THAT INTERVIEW WAS DONE FOR OUR DOCUMENTARY "TESTING OUR SCHOOLS." FOR MORE, VISIT OUR WEBSITE [WWW.PBS.ORG/MERROW](http://WWW.PBS.ORG/MERROW), THAT'S M-E-R-R-O-W.

VALERIE VISCONTI AND CHRISTINE UMALI PRODUCED THIS PODCAST. I'M JOHN MERROW AND WE'LL BE BACK NEXT WEEK WITH ANOTHER PODCAST. THANKS FOR LISTENING.