

Director of The Center for the Study of Testing, Evaluation, and Educational Policy, George Madaus has written widely on the use of tests in public policy. Here he explains how the misuse of tests corrupts the inferences we draw from them, calls for cost-benefit evaluations to determine whether we are really getting our money's worth, and challenges educators to lobby for better ways to fulfill accountability needs.

## On Misuse of Testing: A Conversation with George Madaus

Photograph by Gary Gilbert



**W**hat are your concerns about the tests being used in American schools?

It isn't so much the tests themselves; it's the sanctions or rewards attached to the test results. When important things like graduation, promotion from grade to grade, teacher evaluation, school district certification, and financial support—major aspects of accountability affecting teachers or administrators—depend directly on test scores, it corrupts the process you're trying to improve.

**That's what you call "high-stakes testing."**

That's right. When the stakes are high, people are going to find ways to have test scores go up. That's true of any social indicator; a good example is airplane schedules. When the Federal Aviation Commission started to publish arrival times of airlines—who was late and who wasn't—within six weeks the airlines added a half hour to each schedule. I used to fly into Washington in 60 minutes; now it's scheduled to be an hour-and-a-half flight.

It's the same with test scores. If it's important enough, people are going to find ways to get kids over the hurdle of the tests. The school will look better, but the skill levels will not necessarily be going up. You may have succeeded only in corrupting the inferences you wanted to make from the tests. I see that happening a lot. Instruction is becoming more and more a matter of getting kids ready for multiple-choice tests or teaching them to write simple little paragraphs to get through the state writing exams.

### **How did we get into this situation?**

Historically there have been two trends, one influencing the other. After the Second World War, there seemed to be a change in the way people conceived of education. The prevailing metaphor for schools became the factory or the production model; it may have come out of McNamara's Defense Department mentality of objective-based management and so on.

The other movement, particularly in the '70s, was a demand for accountability. There was a perception that we were putting a lot of money into the schools, but the SAT scores were going down. Admiral Rickover was complaining that Naval recruits couldn't read or write, employers were complaining that kids couldn't fill out applications, so there was a demand to know what we were getting for our money.

That "bottom line" mentality has come to dominate policy making. State and federal legislators passed laws, starting with the National Defense Education Act back in the '50s, requiring testing. There have been any number of bills at the federal level that mandate use of test results in some way to look at outcomes. And that has made schools much more outcome- rather than input-oriented.

**Some say that's a good thing. They argue for outcome-based education, which implies a way to measure outcomes—and of course there are those who argue for measurement-driven instruction.**

Theirs is a value position very different from mine. I just don't think a production orientation is a proper way to conceive of schools. Teaching is not like applying spray paint to an automobile, where the automobile is static and the only thing moving is the paint. You've got live kids and live teachers with all kinds of different backgrounds interacting in social situations with a lot of variables over which the teacher—and the kids—have no control.

**Do you see any benefits at all from a measurement-driven approach?**

## **Measurement-driven instruction takes a lot of important decisions out of teachers' hands.**

In the short run it focuses teachers' attention on objectives; but in the longer run, if teachers, principals, or superintendents start to see those test scores used to evaluate them, or in some other way the results start to really make an impact on them or their students, then the testing will become the curriculum.

The easiest thing in the world for teachers and administrators is to follow the model of measurement-driven instruction—it makes their instruction much easier—but I think it deprofessionalizes teaching and administration. It takes a lot of important decisions out of teachers' hands. Any move to school-based management will have to examine closely the implications of high-stakes measurement-driven instruction for school-based decision making.

**How about the idea of outcome-based education, which may not stress testing per se but does emphasize intended results?**

Well, I'm not going to argue that we shouldn't rationally decide what we want to accomplish, but what tends to happen when you take that outcome-oriented approach is that you focus on things that are easily quantified. A lot of other important things fall by the wayside. When the accountability movement began to take hold, state level people came to realize that they couldn't mandate what teachers were to do in classrooms, but they could mandate a test; and if the test had strong sanctions associated with it, it would modify teachers' behavior.

**In some cases that was purposeful, but in others it may not have been the intent.**

Right. A good example of that is Kentucky, which had what I thought was a perfectly reasonable state testing system. The legislators there did not intend to use test results as an accountability tool; the original idea was to identify kids who needed help. But then the Louisville and Lexington papers, the two biggest papers in the state, decided to rank school districts on the basis of the school testing program. That put enormous pressure on superintendents and principals, who in turn put pressure on teachers. When some of us were asked to evaluate the Kentucky program, we found that those rankings didn't take into consideration a lot of things like demographics; but, most important, they didn't take into consideration the fact that the differences between some rankings weren't real. The point is that the tests were being used in a way different from their original purpose and were perceived as accountability tools. This changed the nature of the testing program.

**Is it accurate to say that part of this is educators' own fault? That we overreact and pay too much attention to test scores when we really should know better?**

I don't think so. Educators are acting in either their own self-interest or

**When you take an outcome-oriented approach, you focus on things that are easily quantified.**

**We need to know how much effort, time, and money are being spent on preparation for high-stakes tests. We also have to see if test gains reflect real gains in the skills in question.**

what they see to be the interests of their students.

**Is it your view, then, that some of the "score increases" we've seen in recent years do not represent real improvement?**

Isn't it interesting that when Cannell published his findings about the "Lake Wobegon" effect, everybody was upset? Now why should they have been upset? The report that scores were going up all across the country should have been greeted as great news—but in their guts, people knew what was going on and just didn't trust the scores.

**I was surprised that people seemed so surprised.**

At any rate, I don't know how much of the improvement is real and how much of it is smoke and mirrors.

**Maybe the answer is better measurement. There's a lot of interest in alternative forms of testing.**

I'm all for that; we need alternatives. But we're locked into the multiple-choice machine-scorable efficiency mode of measurement.

**Let's talk about that. I'll mention several developments that some people consider improvements**

**and ask you to comment. First, what about locally devised criterion-referenced tests such as those used in the Portland and Pittsburgh school systems?**

I know the Pittsburgh system, and I think they're doing an excellent job. I'm not as familiar with Portland, but I do know they have a very large test bank. But, again, if a school system uses only multiple-choice tests and if they are accompanied by important rewards or sanctions, those tests are eventually going to become corrupted too.

What we don't have is good cost-benefit evaluations of what's going on in testing across the country. One of the reasons is that nobody wants to fund that kind of research. Cynically, I think some people are afraid of what might be found. On the other hand, maybe I'm wrong, and real gains are being made that are directly attributable to high-stakes testing programs.

But we haven't looked at the instructional costs of pursuing test-oriented models, such as mastery learning. We don't really know how much time is spent on test preparation, legitimate and illegitimate, in a typical school year. Walt Haney and I have estimated that the dollar cost of simply taking time to administer these tests is about four to five times the cost of purchasing the tests. We need to know how much effort, time, and money are being spent on preparation for high-stakes tests. We also have to see if test gains reflect real gains in the skills in question. Only then can we decide if these expenditures are worth it.

**Another effort to improve things is the development of new reading tests in Illinois and Michigan. They're machine scorable, but their design is intended to better reflect current ideas about what reading is.**

I have great expectations for those tests; I think they have a lot of promise. But the issue isn't the design of the tests as such. It's the decisions, sanctions, and rewards that accompany the tests—these are the problem.

**Would you say the same about performance testing? Connecticut has been a leader in this at the state level, and some of the early National Assessments tried to find out whether students could actually do things, such as sing in tune the first line of "America."**

That way of measuring learning is excellent. In science, for example, and in language, it's really the only way to go. It's not efficient; it's expensive; it doesn't lend itself to mass testing with quick turnaround time—but it's the way to go. To use that approach, though, we've got to trust teachers a lot more, because teachers will have to administer the exercises as well as use the results.

A reasonable approach would be to involve teachers intimately in the development and scoring of the test instruments. Teachers should be in the loop as much as possible. They're in the front lines; they know the kids best.

**Two other improvement efforts I want to ask about are the portfolio system used in Pittsburgh's Arts PROPEL, which Howard Gardner is associated with, and the exhibition idea being tried in Ted Sizer's Essential Schools project.**

Those experiments deserve careful watching. They hold a lot of promise.

**They're the sort of assessment you were talking about—designed pri-**

**Until we come up with alternatives the public will accept, the accountability mentality will be with us.**

## We should argue for testing samples of students, for using multiple indicators of student progress, and for including teacher discernment and judgment in the process.

marily to give feedback to educators working directly with students and planning local programs.

That's the level where I think the information is most valuable and can be used most efficiently. But those approaches don't really lend themselves to accountability models at state and national levels.

**If there were more trust, maybe some of the information could be used to inform the public.**

Yes, trust has to re-emerge. Much of the accountability legislation has come from lack of trust.

**Do you see any signs that may be changing?**

Frankly, no. Until we come up with alternatives the public will accept, the accountability mentality will be with us—for at least another three to five years.

**What, then, should local educators do when they are faced with the printing of test scores in the newspapers, when they are required by the state to give tests they consider invalid, unhelpful?**

As a profession we need to work to lower the stakes on some of these

tests—we've got to get the word out about what is happening. For example, I want people in the Boston area to know, when the state test scores are published in the local papers, what they cost in terms of instructional time and money. We shouldn't oppose testing or accountability, but we should lobby for new ways to accomplish public accountability without large-scale multiple-choice testing of all students at a given grade. We should argue for testing samples of students, for using multiple indicators of student progress, and for including teacher discernment and judgment in the process.

**Business people are used to thinking in terms of cost effectiveness.**

They've got to know the real costs behind getting scores to go up, and they've got to begin to think about whether those improvements are real or not. That's going to take some independent evaluation, but whoever does such evaluations can't work for the state or the school districts. There'd be too much pressure for a positive answer.

We've got to start asking some very specific questions and collecting data to try to answer them. Until that happens, teachers and administrators will continue to be caught in the dilemma. □

1. Early in 1988, John Cannell, a family physician in Beaver, West Virginia, announced that he had canvassed state education departments to find that nearly all reported that their students were scoring above national norms. Fans of folk humorist Garrison Keillor were reminded of his description of the mythical town of Lake Wobegon where "all the children are above average." See Edward B. Fiske, "Standardized Test Scores: Voodoo Statistics?" *New York Times*, February 17, 1988.

**George F. Madaus** is Director, The Center for the Study of Testing, Evaluation, and Educational Policy, and Professor of Education, Boston College, Chestnut Hill, MA 02167. **Ron Brandt** is Executive Editor, ASCD.

**Kathleen A. Butler, Ph.D.**  
invites you to attend her

Fourth Annual  
Summer Institute Series  
on  
**STYLES of  
LEARNING, TEACHING  
and  
THINKING**

Dr. Kathleen A. Butler, noted specialist in learning style and author of the acclaimed book, *Learning and Teaching Style: In Theory and Practice*, invites you to attend these extraordinary programs on style.

**Learning and Teaching Style:  
An Introduction**

Dr. Butler's introduction to the Gregorc Model of Style and her own SDI approach. (Beginner level)  
June 29 - 30, Cape Cod  
August 16 - 17, Wappingers Falls, NY  
August 23 - 24, Great Valley S.D., PA

**Style in the Classroom:  
Dr. Butler's SDI Approach  
to Teaching with Style**

An extremely practical integration of style with a vast array of classroom strategies, including Bloom's Taxonomy. (Intermediate level)  
July 17 - 19, Minneapolis  
July 25 - 27, Marshall, Michigan  
July 31 - August 2, Cape Cod

**Exclusively Staff  
Development!**

A detailed program on Kathleen Butler's approach to style and establishment of an effective staff development program on style within district. (Advanced level)  
August 7 - 11, Cape Cod

**Learning and Teaching Style:  
A Counseling Perspective**

An intensive, exciting program for counselors at all grade levels.  
July 20 - 21, Minneapolis  
August 3 - 4, Cape Cod



For details, write or call:  
Dr. Kathleen Butler  
The Learner's Dimension  
PO Box 6 • Columbia, CT 06237  
(203) 228 - 3786

Copyright © 1989 by the Association for Supervision and Curriculum Development. All rights reserved.