



FAILING THE TEST

BY RICHARD WHITE

The National Assessment of Education Progress (NAEP), a program of the United States Department of Education, has been tracking academic performance for 30 years, as measured by the scores on standardized science, math, and reading tests of students in grades 4, 8, and 12. Following the group's report in August 2000 ("The Nation's Report Card" at <http://nces.ed.gov/nationsreportcard/pubs/main1999/2000469.shtml>), then-Education Secretary Richard Riley noted that reading scores have shown no improvement in recent years. In response, he called on Congress to increase funding for federal reading programs. He also asked parents to help their children by reading to them 30 minutes a day, saying "The trends report finds that reading in the home is down, and that there is a correlation between reading in the home and

[performance on] achievement tests."

Further action is on the horizon under the new administration. Whenever public officials view standardized test results as disappointing, they put in place various incentives—as well as threats—designed to motivate teachers and schools to perform better. As part of his campaign platform, President George W. Bush proposed requiring low-scoring schools to improve or lose federal Title I funding.

Through the Systems Thinking Lens

On the surface, tracking students' test scores seems like a fair way to judge a school's effectiveness. But what are some of the unintended consequences of focusing on testing? "Teaching to the Test" illustrates some of the problems that result from this emphasis. This dynamic is an example of the

"Shifting the Burden" archetypal structure, in which people address a perceived problem by implementing a symptomatic solution that seems to solve the problem over the short term. However, this "quick fix" diverts attention from a more fundamental solution and often results in an addiction to the short-term solution.

In this case, the perceived problem is disappointing test results. The pressure to improve test scores has led to the widespread symptomatic solution of "teaching to the test," or focusing on preparing students specifically for the test-

taking experience (B1). The more fundamental solution is providing students with a rich educational experience that prepares them for the test as a matter of course (B2).

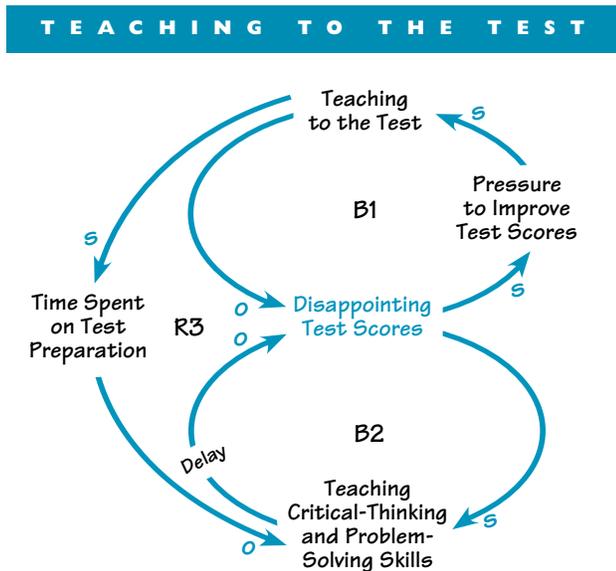
But nurturing students' critical-thinking and problem-solving skills takes time and leads to improvement on standardized tests only after a delay, if at all. Unfortunately, many school administrators believe they need a solution that produces immediate results; their scores must be better next year or the government will punish them. In this way, teaching to the test detracts from educators' ability to build students' thinking and reasoning abilities (R3).

Over time, many school systems, driven by fear, have become addicted to teaching to the test. The obsession with tests and test preparation has even spawned a thriving industry in the private sector, as parents pay to have their kids tutored on test-taking techniques. Thus, the focus of primary and secondary education in the United States is more and more on the tests themselves and less and less on developing children's ability to learn.

Questions for Reflection

- What are some factors that could prevent us from correctly defining the problem to be solved?
- When test scores fall, whom do we blame? On the other hand, when test scores improve, to whom do we generally give credit? What is the role of incentive programs, initiatives, and threats in improving test scores?
- What is the primary goal of our education system, and how effective are standardized tests in measuring our success in achieving that goal?
- What might be the long-term effects of this system on children?

Continued on next page >



The pressure to improve test scores has led to the widespread symptomatic solution of "teaching to the test" (B1). The more fundamental solution is providing students with a rich educational experience that prepares them for the test as a matter of course, which takes time (B2). Teaching to the test detracts from educators' ability to build students' thinking and reasoning abilities (R3).

The Apple's Core

What are some leverage points for breaking this addiction? We can start by examining how the original problem—disappointing test scores—is defined. When we delve more deeply into the subject, we find that errors in thinking have led to ill-considered policies based on invalid conclusions. In fact, teaching to the test is just a symptom of deeper problems; only by addressing these challenges at the system's—and our own—core can we best serve our children's interests.

Enumerative Versus Analytical

Studies. The first mistake we find is that the authors of the NAEP report erred in using enumerative studies when drawing their conclusions.

Enumerative studies make inferences about groups based on measurements taken from samples of those groups. For example, we can infer that the sampled 8th-grade reading score reported for 1999 closely approximates the average score for all 8th graders for that year.

However, by definition, enumerative studies do not provide a basis for comparing results *over time*. Few people—and not even many statisticians—are versed in the practice of *analytical studies*, which brings the necessary statistical rigor to conclusions and predictions derived from such comparisons. By calculating upper and lower control limits to delineate the boundaries of what we can expect from the system, we avoid (1) interpreting an event as an unusual occurrence when it is simply an example of normal variation or (2) interpreting an event as an instance of normal variation when it is actually an unusual occurrence. Typically, only data points outside the control limits warrant attention; those between the limits represent normal variation.

Applying analytical studies reveals that reading scores for all three grades have remained within the control limits over the 30 years of testing. By definition, the system's output is stable—and in order to make any genuine changes, policymakers must transform the system itself. Any improvement efforts that individual

contributors undertake, such as teaching to the test, are futile at best and harmful at worst—incentives, threats, and best intentions notwithstanding.

Correlation Versus Cause-and-Effect. Confusion between the terms “correlation” and “cause-and-effect” often compounds problems of statistical interpretation. If we learn that the number of lifeguard rescues per day at a seaside resort closely correlates to the amount of ice cream concession stands sell, we might conclude that the old admonition about not going in the water too soon after eating is true. But in actuality, increased ice cream consumption and swimmer mishaps show a correlation simply

Demanding improved results from scapegoats trapped within a stable system that they cannot change is a futile, cruel, and unfortunately commonplace occurrence.

because both are effects of the same cause: hot weather. There is no evidence of any *cause-and-effect* relationship between the two.

Our tendency to seek to confirm our beliefs often leads us to assume causal relationships where none may exist, and blinds us to them where they may. For example, Secretary Riley's conclusion that reading at home leads to higher performance on achievement tests may indeed be correct, but he makes a logical error by basing his assumption solely on the correlation between the two. Later, Riley dismisses the report's enumerative evidence that private school students do better on the tests than public school students, saying that such evidence in no way implies that private schools provide a better education than public schools. The evidence should be considered, because enumerative studies are appropriate in this context. Yet Riley can see no possible causal relationship between private school education and higher test scores.

The Fundamental Attribution

Error. In our eagerness to determine causality, we often fall prey to what psychologists call the *fundamental attribution error*. This term describes the act of unjustly attributing the outcome of a complex process to a single contributor. For instance, football fans may blame their team's loss on the place kicker who misses a last-second field goal—distracted from reason by the seductive closeness in time between the kicker's attempt and the moment of loss (the game's end). But they overlook the fact that if the kicker's teammates had performed better, the attempt might have been irrelevant. Similarly, those searching for accountability for disappointing test scores look to obvious contributors, such as teachers, parents, and students, and ignore possible flaws in the system.

Designing Interventions in Dynamic Systems

A lesson from this example is that, when designing interventions in a dynamic system, we must first reexamine our knowledge of the problem. In this case, misunderstandings and flawed assumptions hide the fact that the perceived problem *does not exist*—reading scores have been stable for 30 years. Demanding improved results from scapegoats trapped within a stable system that they cannot change is a futile, cruel, and unfortunately commonplace occurrence.

Thus, our national obsession with standardized test scores—and our well-intentioned but ill-advised reactions to short-term “trends” in them—are scientifically invalid. We are artificially creating winners and losers, thereby making losers of everyone. Most important, we are dedicating our precious time, resources, and emotional investment to wrestling with phantoms, instead of to the worthy goal of raising our children to be joyful and competent learners. ■

Richard White is the founder and president of Orchard Avenue, a consulting firm helping people and organizations achieve greater effectiveness through deeper understanding.