

Teach to the Test?

Most of the problems with testing have one surprising source: cheating by school administrators and teachers.

BY RICHARD P. PHELPS

EVERY YEAR, THE EDUCATION MAGAZINE PHI DELTA *Kappan* hires the Gallup Organization to survey American opinion on the public schools. Though Gallup conducts the poll, education grandees selected by the editors of the *Kappan* write the questions. In 2007 the poll asked, “Will the current emphasis on standardized tests encourage teachers to ‘teach to the tests,’ that is, concentrate on teaching their students to pass the tests rather than teaching the subject, or don’t you think it will have this effect?”

The key to the question, of course, is the “rather than”—the assumption by many critics that test preparation and good teaching are mutually exclusive. In their hands, “teach to the test” has become an epithet. The very existence of content standards linked to standardized tests, in this view, narrows the curriculum and restricts the creativity of teachers—which of course it does, in the sense that teachers in standards-based systems cannot organize their instructional time in any fashion they prefer.

A more subtle critique is that teaching to the test can be good or bad. If curricula are carefully developed by educators and the test is written with curricula in mind, then teaching to the test means teaching students the knowledge and skills we agree they ought

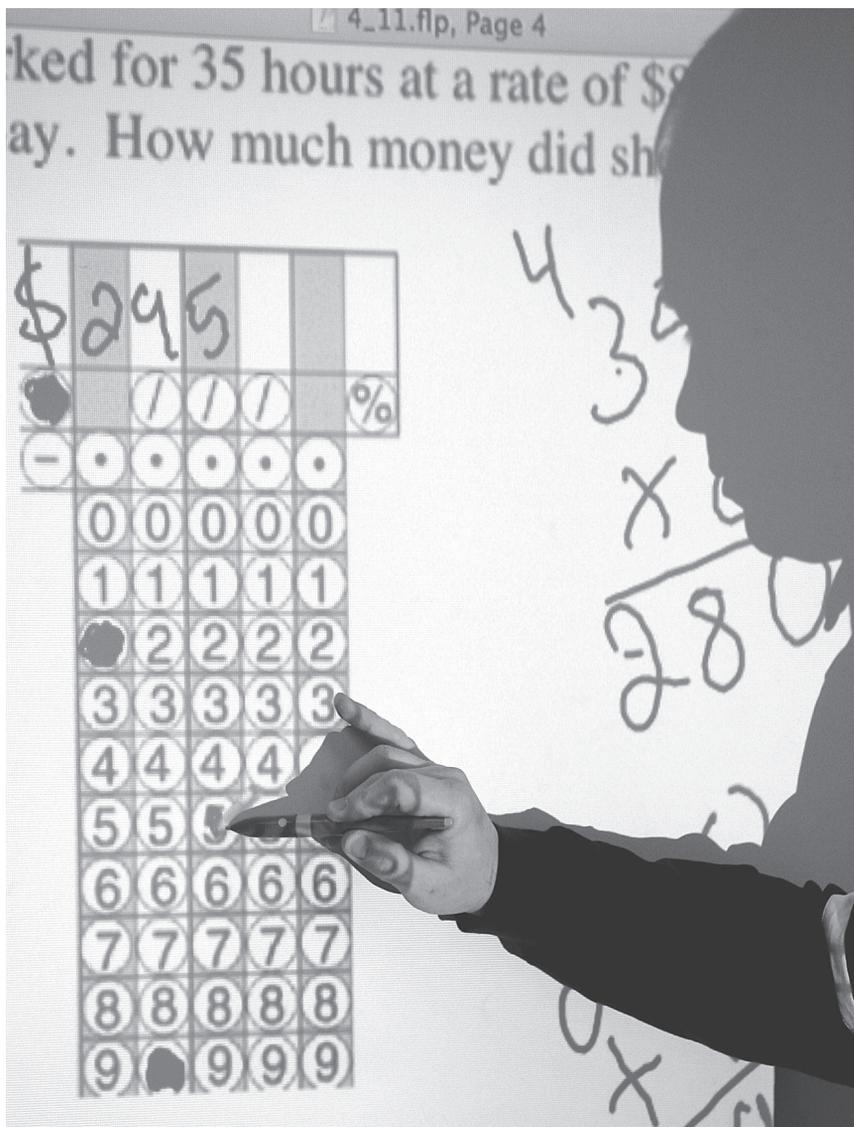
to learn—exactly what our teachers are legally and ethically obligated to do.

Yet there are two senses in which teaching to the test can indeed be harmful: excessive preparation that focuses more on the format of the test and test-taking techniques than on the subject matter, and the reallocation of classroom time from subjects on which students are not tested (often art and physical education) to those on which they are (often reading and mathematics).

The No Child Left Behind Act of 2001, for example, implicitly encourages educators to reallocate classroom time, because it requires testing in only reading and math (in seven grades) and science (in three). Researchers have yet to determine exactly what the effects have been in schools, but NCLB has created a clear incentive for educators who are worried about their schools’ performance to cut back on art, music, and history classes while devoting more time to reading, math, and science. (Since science results are not included in the school accountability calculations under NCLB, however, that subject may also get short shrift.)

What about all the time spent on schooling students in the techniques of test taking—how to fill in answer sheet bubbles, whether to guess or not, what to do when time runs short, and so on? This kind of instruction has been known to eat up weeks, even months, of class time during which students study old examinations or practice test-taking skills. It should

RICHARD P. PHELPS is the author of *Standardized Testing Primer* (2007) and other books about testing and the founder of the online *Nonpartisan Education Review*. He lives in Asheville, North Carolina.



Many teachers drill students extensively in test-taking techniques, but scores don't rise.

occupy less than a day. The firms that write today's standardized tests, such as the Educational Testing Service and CTB/McGraw-Hill, strongly discourage this kind of preparation, correctly arguing that teachers who spend more than a little time familiarizing students with test formats can hurt learning and test performance by neglecting to cover the subject matter itself. (As for the amount of time spent administering the tests, another source of complaints, it is insignificant. The tests required by NCLB, for example, are given once a year and take about an hour each.)

and far more troubling reason why this kind of teaching to the test persists: It sometimes works. And it does so for a very bad reason: Repeated drilling on test questions only works when the items match those on the upcoming test. But if those questions are available to teachers, that means test security has been breached. Someone is cheating.

Test security includes measures ranging from taking effective precautions against divulging any but the broadest foreknowledge of the test's contents to educators and students to guarding against old-fashioned

The evidence from commercial firms that offer preparation for college and graduate school entrance tests such as the SAT and GRE is clear on this point. Most companies, including industry behemoth Kaplan Inc., focus on subject-matter review. However, one firm, The Princeton Review, distinguished itself for years by arguing stridently that students need not master such material to do well. For a fee of several hundred dollars, it would teach test-taking techniques that it promised would increase scores. But dozens of academic studies failed to confirm these claims, and after sustained pressure from better-business groups, The Princeton Review agreed last year to pull the ads in which these assertions were made.

Why do so many teachers persist in extensive test preparation? Partly because they have been misled. But there is a deeper

cheating when students take tests. It requires diligence both in proctoring test administration and in maintaining the “integrity” of test materials. For example, for a paper-and-pencil test, materials must be sealed until the moment test taking begins and students—and no one else—open their test booklets. Students should be the ones to close those booklets, too, with the completed answer sheets inside. Recent cheating scandals around the country, however, indicate how easily and frequently integrity is violated.

Unlike in most other industrialized countries, security for many of our state and local tests is loose. We have teachers administering tests in their own classrooms to their own students, principals distributing and collecting test forms in their own schools. Security may be high outside the schoolhouse door, but inside, too much is left to chance. And, as it turns out, educators are as human as the rest of us; some cheat, and not all manage to keep test materials secure, even when they are not intentionally cheating.

Lax test security has plagued American education for at least a quarter-century. The people in the best position to fix the problem, though, are the same ones who direct our attention instead to the evils of “teaching to the test.” But teaching to the test is not the main problem; it is the main diversion.

It was not always so. In the late 1970s, a group of 10 African-American students who were denied high school diplomas after failing three times to pass Florida’s graduation test sued the state superintendent of education. The plaintiffs claimed that they had had neither adequate nor equal opportunity to master the curriculum on which the test was based. Ultimately, four different courtrooms would host various phases of the trial of *Debra P. v. Turlington* between 1979 and 1984.

“Debra P.” won the case after a study revealed a wide disparity between what was taught in classrooms to meet state curricular standards and the curriculum embedded in the test questions. A federal court ordered the state to stop denying diplomas for at least four years while a new cohort of students worked its way through the revised curriculum at Florida high schools and was tested.

Before *Debra P.*, Florida and most other states that gave graduation tests purchased the exams “off the

shelf” from commercial publishers while leaving responsibility for curricular standards management in the hands of school districts. Given that each state’s standards differed, when they existed at all, commercial tests were based either on an amalgam or, except in Iowa and California, another state’s standards.

Florida’s schools had been teaching state standards, but the standards underlying the graduation test were from somewhere else. *Debra P.* revealed a conundrum: In learning the Florida standards, students were not prepared for the graduation test, but if their teachers taught to the test, students would not learn the official Florida curriculum. The court declared it unfair to deny students a diploma based on their performance on test content they had had no opportunity to master.

Debra P.’s legacy continues to prescribe how high-stakes tests are made. The development of standards-based tests is time consuming and expensive. And the process starts only after the content standards have been set. Today, the standards dog wags the test tail.

Even so, some education insiders rue the effect on instruction. Complete alignment matches the content of the curricular standards, the test, and instruction as well, which means that every teacher in the state must teach the same content in a given grade level and subject area. That notion is anathema to many education professors and others who take the romantic view that each and every teacher is a skilled and creative craftsperson who designs unique instructional plans for unique classrooms. In this view, standardizing instruction “de-skills” teachers. Therefore, teaching to a test must always be wrong.

About the time *Debra P. v. Turlington* was decided, John J. Cannell, a medical resident working in rural Flat Top, West Virginia, read about the claims of local school officials that their children scored above the national average on standardized tests. Skeptical, he investigated further and ultimately discovered that every state that administered nationally normed tests made the same claim, a statistical impossibility. Cannell documented the phenomenon—later called the “Lake Wobegon Effect,” an allusion to radio humorist Garrison Keillor’s fictional hometown where “all the children are above average”—in two lengthy self-published reports.

As often happens after school scandals make the news, policymakers and pundits expressed dismay, wrote opinion pieces, formed committees, and, in due course, forgot about it. Deeper investigations into the issue were left to professional education researchers, the vast majority of whom work as faculty in the nation's colleges of education, where they share a vested interest in defending the status quo.

Cannell correctly identified educators' dishonesty and lax security as the culprits behind the Lake Wobegon Effect. At the time, it was common for states and school districts to purchase standardized tests off the shelf and administer the exams themselves. To reduce costs, schools commonly reused tests year after year. Even if educators did not intentionally cheat, over time

view's now-retracted advertising. Seldom do such critics mention their other reasons for criticizing high-stakes tests: These exams are often externally administered and thus beyond educators' direct control, and the results can be used to judge educators' performance.

Consider two particularly high-stakes tests, the SAT and ACT. A student's score on these tests plays a large role in determining which college he or she attends. But these tests exhibit no score inflation. Indeed, the SAT was re-calibrated in the 1990s because of score *deflation*. The most high-stakes tests of all—occupational licensure tests—likewise show no evidence of score inflation. All of these tests are administered under tight security, and test forms and items are frequently replaced.

THIS YEAR ALONE, educators in Atlanta, Washington, D.C., and Pennsylvania have been caught up in test cheating scandals.

The harmful teaching to the test that Cannell uncovered was, unambiguously, cheating. Is it still practiced today? Probably not widely, but yes. This year, cheating scandals were uncovered in Atlanta, Washington, D.C., and Pennsylvania. The

they became familiar with the test forms and questions and could easily prepare their students for them. When test scores rose over time, administrators and elected officials could claim credit for increased learning.

These were not the high-stakes graduation tests of *Debra P.* Test security was very lax because the tests were given only for diagnostic and monitoring purposes. They "didn't count"—only one of the dozens of state tests Cannell examined carried direct consequences for educators or students. Nevertheless, prominent education researchers, most notably those associated with the federally funded National Center for Research on Evaluation, Standards, and Student Testing (CRESST), at the University of California, Los Angeles, blamed "high stakes" for the test score inflation.

In this line of argument, high stakes drive teachers to teach (successfully) to the test, which results in artificial test score increases. CRESST researchers and others simply ignored the abundant evidence to the contrary—that too much time studying a test format harms students—and, in effect, echoed the claims of The Princeton Re-

800-page Investigation Report on the Atlanta Public Schools named 178 school-based principals, teachers, and other staff who either pressured others to cheat or felt pressured themselves in a "culture of fear, intimidation, and retaliation." The most common illicit activity investigators uncovered was painfully straightforward: Teachers and administrators erased students' incorrect answers and replaced them with correct ones.

In Washington, D.C., school administrators practiced a more elaborate form of score manipulation: the blueprint scam. During a test's development, a contractor typically produces a "blueprint"—a document that matches education standards to the test items written for them. Blueprints show that the draft test items cover all the standards, and in acceptable and consistent proportions. Often they are kept secret along with other test materials until the tests are completed. But some states make their blueprints public, indicating that some standards are meant to be emphasized more than others.

Washington's school authorities go a step further. Each year, they publicly identify a large number of

standards—as many as half the total in some cases—that will not be represented by any test item. Teachers then face a moral dilemma. They are ethically and legally obligated to teach all the standards for their grade level and subject. But the students of their colleagues who do not do so—who teach only the standards they know will be tested—may well perform better on the year-end test. The official record will show the thorough, responsible teacher to be inferior to colleagues who take instructional shortcuts. And in Washington, teachers can be rewarded with pay bonuses or subjected to dismissal on the basis, in large part, of their students' test performance.

Much harmful teaching to the test would be easy to fix: by tightening security, rotating test items and test forms frequently, and squashing sleazy deceits such as Washington's blueprint scam. Test security is more likely to be tight when tests are externally administered, either by computer or by proctors unaffiliated with the schools. If neither approach is possible, test booklets should be made tamper proof, teachers who administer a test should do so in a classroom other than their own, school administrators who handle test materials should do so in a school other than their own, and the materials should arrive just before test time. Neither teachers nor principals can coach students on specific test items in advance if they don't have them in advance. And educators can't change students' wrong answers if they never touch the answer sheets.

The problem is easy to fix, however, only if educators genuinely desire to stop the cheating. Although the fixes are simple and obvious, test security is effectively no better today than it was in Cannell's time. The tests are better, but test security often is not.

The current flawed testing regime puts teachers in a classic “damned if you do, damned if you don't” bind. The only way out, according to many educators, is to eliminate testing, or at least the stakes attached to it. But without standardized tests, there would be no means for members of the public to reliably gauge learning in their schools. We would be totally dependent on what education insiders chose to tell us. Given that most testing critics are education insiders, that may be the point.

The furor over the recent cheating scandals could

lead to real progress on test security reform, but the vested education interests are still trying to deflect attention elsewhere. Earlier this year, the National Research Council released a report that again asserts a causal relationship between high stakes and score inflation and ignores test security's role. The report's proposed solution is to administer new no-stakes “audit tests.” Under the dubious assumption that such no-stakes tests are inherently trustworthy and incorruptible, the resulting score trends would be used to shadow and allegedly verify (or not) the trends in the high-stakes tests. Thus, resources that could be used to bolster the security of the test that counts would be diverted instead toward the development and administration of a test that didn't. Who would administer the new tests? Almost certainly it would be school officials themselves.

A more fundamental worry is that education researchers are now attempting to compromise the Standards for Educational and Psychological Testing, a set of guidelines developed by three national professional organizations for developing and administering tests that the courts use as a semiofficial code of conduct. The education insiders have incorporated their ideas into the draft revision of the standards. In its more than 300 pages, the draft says next to nothing about test security.

We have an opportunity to set things right with an agreement by more than 40 states to embrace the new Common Core State Standards Initiative for kindergarten through grade 12, beginning in 2014. The Common Core is sponsored by the National Governors Association; participation is voluntary. Standards for reading and math have already been agreed upon, and committees are drafting those for science and social studies. The design and administration of the relevant tests are being discussed now. Already, the liveliest debate concerns the lower grades, where many standards—such as those that require students to speak, draw, dance, and build—can be tested only through expensive procedures. If they wish to head off harmful reallocations of classroom time and ensure that “what gets tested is what gets taught,” policymakers will need to spend the extra money. That decision is itself a test of our determination to assure tight security in the new system, and a superior education for American children. ■