

Educators cheating on tests is nothing new; Doing something about it would be

Richard P. Phelps*

In the 1980s, a young medical resident working in a high-poverty region of West Virginia heard local school officials claim that their children scored above the national average on standardized tests. Skeptical, he investigated further and ultimately discovered that every U.S. state administering nationally-normed tests claimed to score above average, a statistical impossibility. The phenomenon was tagged the “Lake Wobegon Effect” after Garrison Keillor’s “News from Lake Wobegon” radio comedy sketch, in which “all the children are above average”.

The West Virginia doctor, John Jacob Cannell, M.D., would move on to New Mexico and, eventually, California, but not before documenting his investigations in two self-published books with titles, “How All Fifty States Are above the National Average” and “How Public Educators Cheat on Standardized Achievement Tests”.

As usually happens after newsworthy school scandals, policy makers and policy commentators expressed disapproval, wrote opinion pieces, formed committees, and, in due course, forgot about it. Deep dives into the topic were left to professional education researchers, the vast majority of whom worked then, as now, as faculty at graduate schools of education, where they shared a vested interest in defending the *status quo*.

Dr. Cannell cited educator dishonesty and lax security in test administrations as the primary culprits in the Lake Wobegon Effect, also known as “test score inflation” or “artificial test score gains”. It is easy to understand why. Back then, it was common for states and school districts to purchase nationally-normed standardized tests “off the shelf” and handle all aspects of test administration themselves. Moreover, to reduce costs, it was common to reuse the same test forms (and test items) year after year. Even if educators did not intentionally cheat, over time they became familiar with the test forms and items and could easily prepare their students for them. With test scores rising over time, administrators and elected officials could claim credit for increasing learning.

Test security was so lax because they were diagnostic and monitoring tests that “didn’t count”—only one of the dozens of state tests Cannell examined was both nationally-normed and “high-stakes”—involving direct consequences for the educators or students involved.

Regardless the fact that there were no stakes attached to Cannell’s tests, however, prominent education researchers blamed “high stakes” for the test-score inflation he found. Cannell had exhorted the nation to pay attention to a serious problem of educator dishonesty and lax test security, but education insiders co-opted his discovery and turned it to their own advantage.

“There are many reasons for the Lake Wobegon Effect, most of which are less sinister than those emphasized by Cannell,” said the co-director of a federally-funded research center on educational testing. After Dr. Cannell left the debate and went on to practice medicine, this federally-funded education professor and his colleagues would repeat the mantra many times—high stakes, not lax security, cause test-score inflation—in dozens of reports published both by their center and by the National Research Council, whose educational testing research function they co-opted.

It is most astonishing that they stick with the notion because it is so obviously wrong. The SAT and ACT are tests with stakes—one’s score on either helps determine which college one attends. But, they show no evidence of test-score inflation. (Indeed, the SAT was re-centered in the 1990s because of score *deflation*.) The most high-stakes tests of all—occupational licensure tests—show no evidence of test-score inflation. Both licensure tests and the SAT and ACT, however, are administered with tight security and ample test form and item rotation.

Spot the causal factor

| | High security (external administration) | Lax security (internal administration) |
|---------------|---|---|
| High stakes | No test-score inflation e.g., SAT, ACT, licensure examinations | Test-score inflation possible e.g., some internally administered district and state examinations |
| No/Low stakes | No test-score inflation e.g., NAEP, other externally administered examinations | Test-score inflation possible e.g., some internally administered district and state examinations, such as those Cannell investigated |

Current test cheating scandals in Washington, DC and Atlanta once again draw attention to a serious problem, and this time there is no doubt that stakes *are* involved. With the No Child Left Behind Act, schools can be rewarded with cash, or punished through reconstitution or closure, depending on their students' test scores. So, as they have now for over two decades, most educators blame the stakes and alleged undue pressure that ensues for the cheating. Their recommendation: drop the stakes and reduce the amount of testing.

Meanwhile, twenty years after J. J. Cannell first showed us how lax security corrupts test scores, regardless the stakes, test security remains cavalierly 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 of them cheat and not all of them manage to keep test materials secure, even when they aren't cheating.

The furor over educator cheating scandals in Atlanta and Washington, DC could lead to real progress on test security reform so long as the vested interests do not continue to control the debate and determine the policy outcome as they have with Dr. Cannell's legacy.

And, they are trying to. The National Research Council releases a report this summer that, again, asserts a causal relationship between stakes and test-score inflation and ignores test security's role. Their solution to the problem is not to increase security, but to administer no-stakes "audit tests" to shadow the high-stakes test administration over time, under the presumption that any no-stakes test's scores are trustworthy and incorruptible. Thus, resources that could be used to bolster the security of the test that counts will be diverted instead toward the development and administration of a test that doesn't. That other test that doesn't count will almost certainly be administered with little security by school officials themselves.

With any high-stakes test subject to audit by any low-stakes test, its perceived quality will be determined entirely by the low-stakes test. Indeed, those who oppose high-stakes testing could add an easily manipulated and unmonitored low-stakes test and tailor it to discredit score gains on their jurisdiction's externally-mandated and monitored high-stakes test.

Even worse, the same education researchers who have co-opted federally-funded and National Research Council work on educational testing are attempting to compromise the *Standards for Educational and Psychological Testing* which, after more than a decade is currently being revised. The *Standards* is a set of guidelines for developing and administering tests. In the absence of any good alternative it has been used by the courts as a semi-official code of conduct. Thus, it has profound impact beyond the boundaries of the relatively tiny

community of testing professionals. The education insiders have incorporated into the draft revision of the *Standards* their notion that stakes, not lax security, cause test-score inflation and audit tests are the way to control it. Meanwhile, in over 300 pages, the draft *Standards* says next to nothing about test security.

The most fundamental issues in these school scandals are neither cheating, nor pressure, nor testing; they are power and control. Standardized test scores will be trustworthy if responsible *external* authorities control their administration. It is that simple. Leave control of testing, or “audit testing”, to school administrators themselves, and wide-scale institutionalized cheating on educational tests will be with us forever.

* Richard P. Phelps is co-author and editor of *Correcting Fallacies about Educational and Psychological Testing* (American Psychological Association, 2008/2009).