

# EDUCATION NEWS

Search

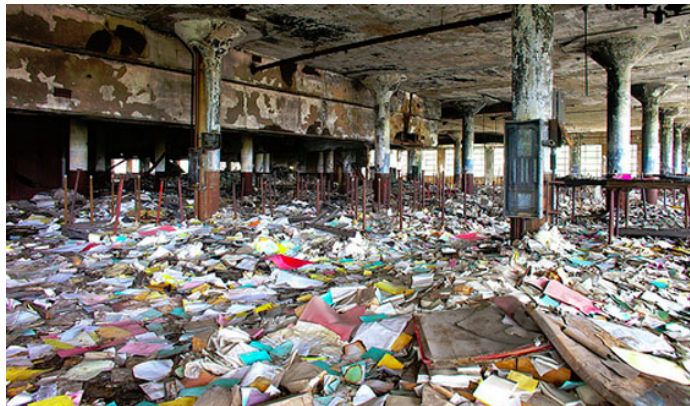
EDUCATION POLICY TECHNOLOGY HIGHER EDUCATION ONLINE SCHOOLS PARENTING INTERNATIONAL / UK

CAREER INDEX

**CAREER INDEX**

# Richard P. Phelps: The School-Test Publisher Complex

Facebook Tweet Delicious Digg StumbleUpon Reddit



by Richard P. Phelps

Within several months, the most important document in US education testing—the Standards for Educational and Psychological Testing—will incorporate the conclusions of biased, irreparably flawed research that favors education’s vested interests. School districts and taxpayers will be compelled to pay for the

TUESDAY  
10 2, 2012

RICHARD P. PHELPS

FILED UNDER

RICHARD PHELPS TESTING

PRINT

EMAIL

## CAREER INDEX

Plan your career as an educator using our online database information.

SUBJECT

STATE

STATE + SUBJECT

CITY + SUBJECT

### ON TWITTER

@EdNews

Coding bootcamp offering more opportunities to and higher salaries  
[#edtech](https://t.co/RC#edtech) [#education](https://t.co/RC#education) [#coding](https://t.co/RC#coding)

21  
A new, unique partnership between Khan Academy and the San Francisco 49ers promotes [#education](https://t.co/11v8MQadcr) [http://11v8MQadcr](https://t.co/11v8MQadcr)



administration of more tests, perhaps twice as many in some areas. But, these new tests will not be used for any of the proven benefits of testing, such as feedback or motivation. Their only purpose will be to “audit” other, already-existing tests.

Why do current tests need “auditing” you ask?

Allegedly, scores and score trends on standardized tests with consequences, or “stakes”, can never be trusted and need to be verified by those from parallel “no stakes” tests. Presumably, scores from no-stakes tests, no matter how administered and no matter who administers them, are as trustworthy as a pug-nosed Pinocchio.

The notion reminds me of the Will Smith–Jon Voight film *Enemy of the State*, in which corrupt politicians and federal intelligence agents misuse their power to monitor their fellow citizens for mutual self-aggrandizement. After the miscreants’ criminal activity is exposed, officials promise to “monitor the monitors”, apparently within the same institutional structures that harbored the original malfeasance. To that announcement, the Regina King character in the film replies “Well, who’s gonna monitor the monitors of the monitors?”

How do those who subscribe to the high-stakes untrustworthy/no-stakes trustworthy belief explain the interaction? They argue that scores and score trends on high-stakes tests cannot be trusted because the stakes induce educators to “teach to the test”, which artificially raises scores over time, producing “test-score inflation.

The alleged empirical support for the proposition is not just flawed it is astonishingly slim, as even the primary advocate of the theory, Daniel Koretz, admits. What Koretz consistently neglects to divulge is that the empirical support disproving the theory is abundant, and far more rigorous than that which he cites. Dozens of experimental and case studies of teaching-to-the-test’s effect have failed to affirm CRESST’s and Koretz’s

---

RT @VsauceTw  
2 of Hyper Hist  
never see this b  
@googlearts [w  
https://t.co/8U  
https://t.co/...

---

## ON FACEBOOK

---

 Like 6.4K

No related posts

## RECENT NEWS

---

FRIDAY, SEP 30, 2016

Judge Orders 1  
Access for Ohio  
Transgender S

THURSDAY, SEP 29, 2016

Holcomb Prop  
Build on Pence  
Reforms in Inc

THURSDAY, SEP 29, 2016

Texas Board of  
Continues Batt  
Evolution in B

## POPULAR NEWS

---



**NUMBER OF  
HOMESCHOOLERS  
GROWING NATIONWIDE**

92

[READ MORE](#)



claims. Instead, they have shown that more than a smidgen of time spent on test-taking skills and test format familiarity reduces the amount of time devoted to teaching the subject matter on which students will be tested, and test scores often drop. Students whose teachers sacrifice essential subject matter lessons for test-taking drills do their students no favors, and may harm them.

Moreover, no-stakes tests are just as prone to test score inflation as high-stakes tests. All but one of the dozens of compromised tests in John J. Cannell's celebrated "Lake Wobegon Effect" studies were no-stakes tests administered without test security protocols. Their positive score trends were expropriated in some cases by education administrators for political boasting. High-stakes tests may come with incentives to cheat, but they tend to be administered with tight security. Most no-stakes tests are administered with lax or no security.

Educators are human beings who respond to a wide variety of incentives. The external pressure induced by high-stakes is but one of many. Few outsiders pay much attention to how no-stakes tests are administered, leaving educators free to manipulate or neglect various aspects as indeed they are "incentivized" to do. For example, schools can save considerable money and time by simply reusing the same test form year after year.

Many outside observers seem unaware that no-stakes tests are generally purchased by education officials "off the shelf" and administered when and how they please. Once purchased, education officials can keep the test forms as long as they like and use them as they like, and even do the scoring themselves. After the point of purchase, test publishers may no longer be involved at all. Or, they may be involved in insidious ways. During his investigation of test-score fraud in the 1980s, J.J. Cannell telephoned test publishers while pretending to be a local school official and bluntly inquired as to how



**ALICE – IN WAL**

[READ MORE](#)



**MATHEMATICS |  
BEING OUTWITT  
STUPIDITY 80**

[READ MORE](#)

he could manipulate test administrations to artificially raise scores. Shockingly, he found company sales representatives unhesitating in their advice, such as, for example, to purchase older, less expensive, but already widely-used test forms and repeatedly re-use them.

Teaching to the test boosts scores only when educators know the content of the test in advance. With secure tests, educators do not know test content in advance. Teaching to the test, then, is a problem only when security is lax. And, if test security (or, “the integrity of test materials”) is the fundamental problem, as a vigilant taxpayer might suggest, why not secure current tests rather than throw money at meaningless, ambiguous, and expensive parallel audit tests?

What gives a weak theory its oomph is the massive marketing power of the two small research groups who advocate for it: the federally-funded Center for Research on Evaluation, Standards, and Student Testing (CRESST) with which Koretz has long been affiliated, and the equally small and well-funded group of Republican Party-affiliated education policy advisors. That CRESST pushes the theory is understandable—discrediting externally-mandated high-stakes tests and promoting internally-controlled no-stakes testing serves education’s vested interests. That the Republican policy advisors push the theory is befuddling, given both that it is wrong and that it is antithetical to their party’s stated interests. It has served well, however, to advance the careers of those involved as only ascription to education establishment orthodoxy can.

In his last major speech as US president, Dwight Eisenhower warned his fellow citizens to be wary of the “military-industrial complex”, the cozy, mutually-beneficial relationship between private arms manufacturers and public officials, be they congresspersons with weapons makers in their districts or military brass eager to deploy those weapons and win combat medals. The mutually-reinforcing nature of

the relationship, and the absence of powerful countervailing political forces, could draw the country into military engagements inimical to its fundamental interests.

The society of education testing experts—education psychometricians—is one of the professional world’s coziest. Their graduate training requires fastidiousness and mastery of a particular grab-bag of statistical techniques. After graduation, employment is available with private or non-profit test publishers, in academe, managing federal, state, or local district testing programs, or in consulting. Their paths frequently cross at their several annual professional meetings, while working together on test development contracts or serving on state or local test advisory committees, or as expert witnesses in test-related court cases. Movement in, out, and in between institutions and market sectors is frequent. They know each other, know that they will be working with each other frequently throughout their long careers, and so “go along to get along”. Disruptive voices are unwelcome.

Why does the test policy section of the new Standards—chapter 13—ignore the vast majority of relevant research and support CRESST doctrine? One possible reason: CRESST researchers wrote it.

Another section of the new Standards, the educational testing section—chapter 12—is exceptionally long, and full to bursting with cautions for test developers and new proposals to check, diagnose, and monitor every test aspect during every test phase, from initial development to the last administration, scoring, and reporting. Particularly in the case of tests with consequences, how could it hurt to err on the cautious side? Incidentally, chapter 12 was written by a vice-president at the world’s largest test publisher, a test publisher that offers, for a price, services to check, diagnose, and monitor every test aspect during every test phase, from initial development to the last administration, scoring, and reporting.

Despite its length, chapter 12 lacks two characteristics. The first is any challenge to the biased research and anti-testing advocacy one finds in the following, test policy chapter 13. Indeed, in general it is rare to witness a commercial test publisher challenge biased research and anti-testing advocacy when the researchers and advocates responsible are highly-placed and politically powerful. After all, education's vested interests are their customers. Besides, they do not necessarily lose business as a consequence of the attacks against testing. They sell many of the fixes and substitutes for testing's real and imagined failings.

The second characteristic conspicuously absent from the educational testing chapter and, indeed, from the new Standards in its entirety, is any detail on test security procedures (or assuring the integrity of test materials) before, during, and after test administrations. Perhaps this is because, like other education psychometricians, the author considers such not to be part of his job. His company sells tests and test services to practicing educators; administering the tests is their responsibility. And, what successful wholesaler tells his retail customers how to run their business?

See also:

[Educators Cheating on Tests is Nothing New; Doing Something About It Would Be](#)

[Extended Comments on the Draft Standards for Educational & Psychological Testing \(but, in particular, draft chapters 9, 12, & 13\)](#)

Richard P. Phelps is the author of [Standardized Testing Primer](#) (2007) and other books about testing and the founder of the [Nonpartisan Education Review](#). He lives in Asheville, North Carolina.

 [Bio](#)

 [Latest Posts](#)

 **Richard P. Phelps**

**0 Comments**

Sort by **Oldest**



Add a comment...

 Facebook Comments Plugin

---

[Privacy Policy](#) [Education](#)