

A Decade in the Paranoid World of US Education Research,

Part 1: The Scoldings

by Richard P. Phelps

The following could be considered one case study—of one person's eyewitness account of a career in education research—or ten case studies of encounters with information suppression efforts.

Either way, the primary conclusion to be drawn is that information suppression in education research is wide-ranging and pervasive. Some of it is well organized; much of it is widely dispersed among a multitude of volunteer censors who feel called to do their part to control the narratives of education research.

Many professional educators feel it right and proper to control the flow of information: denying the public some even though accurate and disseminating some that at best is only partially true. They seem to feel that the education enterprise is their property, despite the fact that all citizens pay for it and most students are other people's children.

This article describes scoldings I endured during one decade working in the education research "industry," for contract research firms serving a range of education organizations, from local school districts to international agencies, and for educational test development firms. Ten scoldings are described one-by-one in some detail in the Appendix. Immediately below, I summarize the scoldings in three ways: by their common characteristics; their effects (i.e., "lessons learned"); and the implications of those effects.

Scoldings represent only a small proportion of the information suppression efforts I, and surely many others, have encountered in our careers. Far more numerous have been less overtly emotional: shunning and ostracism; rejections from potential employers, scholarly journals, or professional meetings for reasons unrelated to qualifications or work quality; and deliberate misrepresentations of our work. (See, for example, Phelps 2015.)

Some readers, of course, will notice that one common element to all the scoldings is yours truly. Couldn't it be something about me that caused the adversarial reactions? My career's chief organizational nemesis, the Center for Research in Educational Standards and Student Testing (CRESST) has apparently said as much. If my sources are accurate, one theme of their quarter-century long character assassination has been that I am too emotional.

I confess to one reaction-in-kind during the decade under review here. After Scolder #1 asserted that I was too unimportant to be worth anyone's time, I countered by calling him pompous. That's it. I endured the other nine scoldings—some of them an hour long—with equanimity. I neither countered emotion with emotion, accusation with accusation, raised voice with raised voice, nor insult with insult. But, neither did I capitulate. I felt that what I was doing was right and proper and stood my ground.

Perhaps I can be rightly accused of naïveté, though. Unlike most of the scolders, I had never attended a graduate school of education, and so did not share some of their common professional socialization.

Introduction

Read widely enough in education research and one is sure to notice oftentimes starkly different beliefs about what is true, even among experts. One prominent disagreement, for example, exists between cognitive scientists—typically university psychology professors—and US education school professors.

For their part, cognitive psychologists wonder how theories of learning that they feel they have thoroughly discredited through decades of experiments remain not only popular, but dominant, in US graduate schools of education and, consequently, among teachers and school administrators. Among those theories: learning styles, the learning pyramid, whole language, discovery learning, and constructivism in general.

Explanations for the endurance of scientifically unsupported folk beliefs typically include one of two:

The enduring appeal of romanticism that one can trace back to Jean-Jacques Rousseau in the 18th century, and his conviction that learning is best left to be a natural process, which the imposition of artificial institutions, such as schools, teachers, and classrooms, can hinder, impede, and distort.

An apparent correlation between education schools' egalitarian preference and the learning theories they advocate. Students who act out in class, for example,

may simply be "kinesthetic learners," so their "learning style" or "multiple intelligence" should be accommodated, not suppressed.

These explanations may be helpful for understanding the motivations of educators who continue to cultivate disproven theories. But, they do not explain how the cultivation works in practice and, just as important, how rival evidence, such as that of cognitive psychologists, remains unconsidered.

That explanation involves pervasive information suppression in US education research. The rival information is generally not included in education school curricula (Kramer 1991, Schalin 2019). Or, it may be presented as wrong, perhaps as propaganda of the ill intentioned.

In the Appendix, I describe some events that occurred over the course of a decade in my education research career. I doubt that my experience was unique, though I may be more stubborn and may have stuck it out for longer than most.

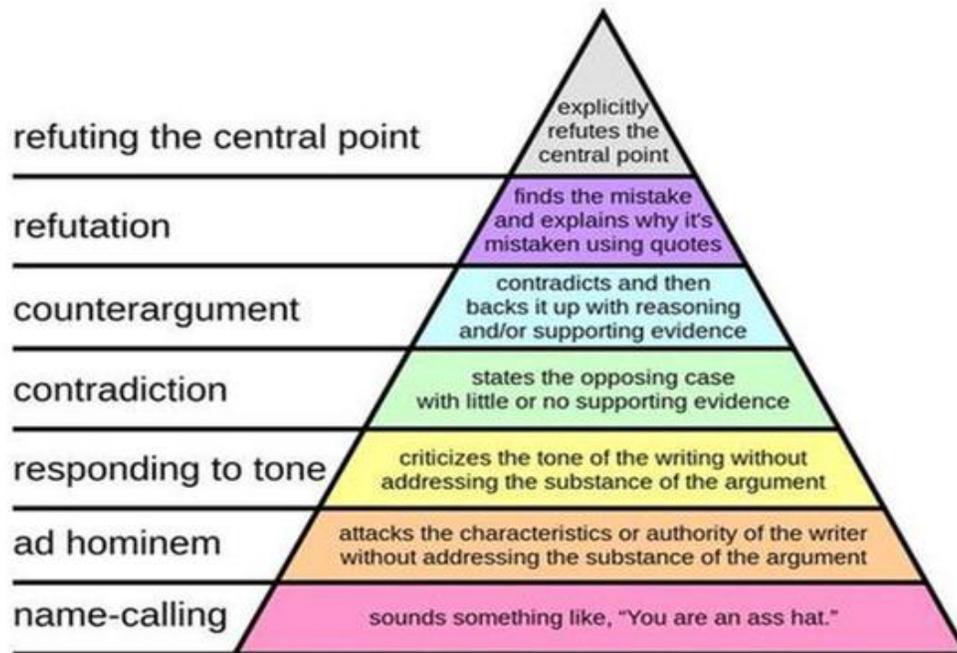
In the end, I, too, would give up, and leave the public education research industry, as so many others have. Weary of the antagonism, the constant threat to my livelihood, and the constantly looming uncertainty, of not knowing where the next attack might come from or what might trigger it.

I was scolded by the head of a firm specializing in industrial-organizational psychology for defending the honorable research tradition of industrial-organizational psychology. I was scolded by managers at educational testing firms for defending the honorable research tradition of educational testing. Indeed, I found it impossible to know when and from where the next scolding—and threat to my career—might emanate.

Characteristics of the Scoldings

For those disinclined to read the Appendix, I summarize here.

Paul Graham's Hierarchy of Disagreement offers a useful reference. He ranks from bottom to top seven commonly used tactics in disagreements.



The Hierarchy of Disagreement, by Paul Graham.

1. Of the seven levels of disagreement in Paul Graham's hierarchy, the scoldings I endured never rose above the first three—name-calling, *ad hominem*, and tone policing. There were no substantive disagreements, no contradictions, counterarguments, or refutations. Indeed, the scolders sometimes agreed with the substance of my evidence and arguments in private. It's the public exposure of that evidence and those arguments that they wished to quash.

2. Outrage was common. Scolders were flush in the face, animated, gesticulating, and angry. It is not just that they did not like what I said or wrote; they felt that I had crossed some boundary of civilized discourse. They seemed to feel that I was *morally* wrong.

The emotion seemed to justify the associated action—rejection or dismissal. What I had said or written was outrageous, unacceptable, impermissible, unconscionable. No matter if what I had said was simply, factually, or obviously correct. No matter that the scolders were professional educators, working in a profession that should value open-mindedness, tolerate disagreement, and welcome a diversity of viewpoints. (See, for example, Haight 2017.)

3. The knee-jerk tendency to suppress unwanted information seemed to require only the tiniest of sparks to touch off highly explosive reactions. Furthermore, emotional outrage might be just the most immediate expression. A hateful effort to damage the career of "the other" might follow. Some scolders could fester

grudges, manifest in months-, years-, or even a decades-long character assassination.

4. The scolders shared much in common with hecklers—audience members who attempt to silence public speakers by shouting over them. A "heckler's veto" occurs when one or a few outraged individuals believe themselves justified in preventing a multitude of others from hearing a speaker, and considering for themselves what the speaker has to say (Olson 2017).

The chapter I wrote for a book to be published by arguably the world's premier authority on standardized testing had been shaped and revised by several internal and external editors and independent reviewers over the course of a couple of years. But Scolder #9 was certain that she knew better and was right to demand that the chapter be scrapped.

Editors of peer-reviewed journals—Scolders #5 and #8—decided not to send my submitted manuscripts out to reviewers. They believed that they knew better and were justified in bypassing the normal manuscript review process.

An article had been published in one of the peer-reviewed flagship journals of Scolder #2's primary professional association, and the organization's leadership had invited me to present the work at their next annual meeting. But, Scolder #2 knew better; the article was beyond the pale of acceptable discourse.

5. Scolders' demands could be contradictory. Scolder #4 told me to be honest and forthright but spent about an hour strongly suggesting I be otherwise. Scolder #3 labeled my soft-spoken, completely impersonal disagreement an "*ad hominem* attack." Scolder #9 called my empirically based research disagreements "personal attacks." Scolder #7 characterized my research disagreements as "censorship."

One can critique others' research without naming names, of course. When I have done it, however, I have been criticized for that, too, for example in the quote in Scolding #8. Not naming names can seem cowardly or suspicious. How can a reader check what you say about someone's research if you do not identify him or her? Why should a reader respect your point of view if you are not even willing to identify those with whom you disagree, thereby giving them an opportunity to rebut?

It's a Catch-22. Naming names is a "personal attack." But, not naming names is cowardly and suspicious.

6. Scoldings typically occurred in private, which, to me at least, seems cowardly. If there's nothing wrong with what one has to say, why not say it openly; have a public debate? Except that public debate may be the last thing they wanted. Privacy offers deniability later?

Ironically, I could be scolded for an allegedly poor "tone" in my public writing by persons exhibiting a surfeit of poor tone themselves behind closed doors. But, I doubt that there existed any tone on my part that would have satisfied them. The problem was that I correctly and factually contradicted research that they wished to protect from scrutiny.

In a further irony, scolders wished to suppress my evidence that other researchers were suppressing evidence (e.g., by declaring a large research literature nonexistent).

7. Punishment could be hugely disproportional to the alleged crime. Scolders were willing to put my livelihood at risk under the tenuous assumption that friends or power brokers in the profession might be personally piqued because someone disagreed with them. Someone's possible feeling of offence outweighed the profession's purported search for truth, its ethical standards, and someone's career.

8. Though their motivations for scolding were posed as altruistic (i.e., they were defending the integrity or image of something or other), they seemed to be largely personal. Professionals at testing firms were not so much offended by my conclusions about educational testing, which were supportive, as they were that I publicly disagreed with other researchers in the profession with the power to assign them to prestigious committees, nominate them for awards, and include them on lucrative contracts. Scolders could be more protective of their professional relationships than they were of research ethics in general or even the reputation of the organization that paid their salaries.

9. Though Scolder #1 expressed it most explicitly, condescension imbued other scoldings as well. I was simply not an important enough person to doubt the wisdom of the high and mighty. Of course, if dissenters are always denied a hearing, they can never reach the same status level as those they disagree with. One might argue, for example, that those most in need of tenure protection in faculties of education—the dissenters—never get that far.

10. Even while claiming to be protective of research truth and the integrity of their products and services, many education research organizations operate primarily as businesses, as scoldings #4 and #10 illustrate most directly. As such, pleasing clients who pay them, or not offending "strategic partners," can be considered more important than adherence to annoying research standards and ethics.

11. The scolders were not obscure cranks; some ranked among the profession's top leadership.

12. While enablers, such as the scolders, worked diligently to ensure that powers-that-be in the profession were protected from the slightest possible offence or disagreement, those powers were free to say and write whatever they pleased. (See, for example, Phelps 2016.)

I do not wish to paint with too broad a brush, however. Most individuals in education research do not behave as the scolders have. That should be obvious: after all, it was others within the profession who published my writing that the scolders tried to suppress.

Indeed, in the middle of the featured decade I worked at an educational testing firm unconcerned with what I wrote outside of work. I did not associate any of my writing with them or list them as an affiliation. I left them alone and they left me alone. And, they suffered not at all as a consequence of what I wrote outside of work coincident with my period of employment there. The same goes for my long stints of employment before and after the featured decade.

Which only serves to show that the crises the outraged scolders alleged to be imminent due to my perceived outspokenness existed wholly inside their heads. Ultimately, despite the scolders' rhetoric of impending doom should my writing be made public, I believe that they were simply trying to suppress information they personally disliked, or disliked being made public. They felt compelled to protect a professional image, one they considered so insecure that any public disagreement might tarnish it.

Even if information suppression is not ubiquitous among education researchers, however, it is widespread enough to keep a lid on public debate in certain topics.

I may have been more courageous (or, reckless) than others in the education industry, but dissent stifling worked on me, too. After all, I needed to make a living just like anyone else. I cancelled my participation in organizations that seemed to be considered heretical by some education insiders. I removed some publications from my C.V. I reduced my volume of writing on education policy. I stopped submitting any manuscripts to mainstream education research journals. I spoke in public less. I adopted pseudonyms for the little I did write for publication.

Finally, I gave up entirely. After the paranoid decade was over, I purposely selected employment unrelated to public education.

After I was hired, the organization's director told me that one of my job references had said that I insisted on telling the truth in research reports, and he reassured me that I would never have to lie while working for them.

But, the referrer was wrong. I would have lied if managers had told me that they wanted me to. After all, if an organization is paying my salary, I am obliged to do the work they give me.

In my experience, however, some managers who wish to publish dishonest research want their underlings to author the reports, perhaps to evade responsibility themselves. While I was never told to lie or write a certain research result, I was sometimes expected to do both. And, when I didn't, when I told the truth as they had told me to do but did not want me to do, outrage might erupt. Once in outrage mode, with emotion overruling reason and fairness, they could summon the rationale to treat me unreasonably.

I can now understand why so many researchers who evaluate educational programs early in their careers end up studying different topics or leaving the profession altogether.¹ Those finding positive outcomes to disfavored programs, such as those with externally administered high-stakes testing, typically do not rise very highly in the education research world. Those finding negative outcomes to disfavored programs are more likely to find career success, even if their research methods may be somewhat sketchy.

I can now understand why so many psychometricians' careers drift toward the occupational licensing world, and away from educational testing. The work is just as well paid, less aggravating, and demands fewer compromises.

Lessons Learned

Bias and fraud saturate US education research, just as censorship and suppression clog its dissemination. Spinning the situation positively, one might say that some prevailing beliefs are aspirational. Leading education researchers desire certain research conclusions to be true because they consider them morally preferable. Research reaching desirable conclusions is celebrated, and can be fiercely defended. Likewise, research reaching undesirable conclusions may be ignored or suppressed, or fiercely attacked.

1. One is free to conduct honest, objective, and divergent research, and even get it published somewhere, but it likely will not be widely disseminated or receive much attention from education journalists or policymakers if it contradicts the profession's aspirational beliefs. Indeed, it will probably be ignored by the information gatekeepers who matter, or even declared nonexistent with statements like "there exists no research (or no 'quality' or 'rigorous' research) reaching (your work's) conclusions." One's work will sink into the vast pool of

¹ For example, in my search of a century's worth of research on the effect of testing on student achievement, I found a dozen or so excellent graduate student theses that arrived at the politically incorrect conclusion on standardized testing—that its benefits can be substantial and outweigh its drawbacks. Typically, those students either continued on to obscure careers inside the profession, or left the profession entirely.

millions of inconsequential research studies, its obscurity serving to reinforce an assumption that it must be of poor quality and, thus, not worth other scholars' valuable time.

2. Should your research expose bias and fraud in popular research, however, you will pay a price. Your career will be stunted, perhaps even ruined. If your work is not successfully suppressed, it likely will be misrepresented. Your personal character may well be smeared. You will earn less money, publish less, win fewer awards, receive fewer invitations, and retain fewer friendships within the profession.

3. Might some of the scolders have thought that my disagreements were fine, just not fine for me to be making them? Were they assuming that someone elsewhere should or would manage the critiques? If so, who? I doubt that I am the only education researcher in the country who has been told to shut up. Most psychometricians working at firms that develop educational tests walk on eggshells these days, afraid to say anything that might possibly offend anyone. They must sign gag orders, meaning the only voices allowed to speak to outsiders reside in the public relations office, from whence only the most saccharine declarations escape.

4. No matter how wrong it might be, one cannot criticize—without adverse consequences—the work of those affiliated with the longtime federally funded Center for Research on Educational Standards and Student Testing (CRESST).² This is true despite, and perhaps because of, the fact that they habitually cite their own research and that of their close allies to the exclusion of all the rest. Cherry-picking research this way—ignoring or declaring nonexistent the vast majority of research evidence in favor of one's own—can "prove" whatever one wishes to prove.³

The group is simply too powerful, too influential, and directs too many research grant dollars and professional appointments. In effect, they have dominated the work on educational testing policy at all of the following institutions:

² Though headquartered at UCLA, the CRESST consortium's organizational members have included the Rand Corporation, Boston College, and the Universities of Colorado, Southern California, and Pittsburgh. Its loyal alumni work at Harvard and Stanford Universities and elsewhere.

³ The two most profound accomplishments of CRESST and its allies: employing highly misleading research, and some condescension, to defuse the infamous Lake Wobegon Effect scandal in the 1980s (and, thus, support continuing widespread lax test security practices in US schools); and convincing multitudes that research evidence to support the use of high-stakes education tests does not exist (or is not any good).

The National Research Council;⁴
 The National Academy of Education;⁵
 The International Academy of Education;⁶
 The World Bank;⁷
 The Organisation for Economic Co-operation and Development (OECD);⁸
 The American Academy of Social and Political Science;⁹
 The American Educational Research Association (AERA); and
 The National Council on Measurement in Education (NCME).

Plus, they seem to easily win funds to publish their research (and ignore or discredit others') from the US federal government and a large number of think tanks and foundations.

Few education researchers stand up to them. Many curry their favor.

Consequently, dozens of erroneous research claims made by CRESST group members over the years remain unchallenged in mainstream communications. They can easily be proven wrong simply by consulting the larger research literature that CRESST loyalists ignore or dismiss.¹⁰

⁴ <https://www.nap.edu/author/BOTA/division-of-behavioral-and-social-sciences-and-education/board-on-testing-and-assessment>

⁵ <https://naeducation.org/our-members/>

⁶ <http://www.iaoed.org/>

⁷ <https://blogs.worldbank.org/team/marguerite-clarke>

⁸ <https://www.tandfonline.com/doi/full/10.1080/0969594X.2014.921091#.VEvBOIcQaVI>

⁹ <https://journals.sagepub.com/toc/anna/683/1>

¹⁰ Among the falsehoods: a) There is no, almost no, or not any good research: showing any benefits to high-stakes testing; on test preparation or test coaching; on testing costs; comparing testing's costs to its benefits; on the volume of testing over time; on postsecondary testing; evaluating test-based reforms; on the design of accountability systems; on educators' response to testing; on student performance at the end of high school; about parent and policymakers feelings about testing; on the effects of testing on teaching; on student performance incentive programs; on methods for checking the alignment between standards and tests; showing a relationship between high-stakes testing and economic growth; on how tests are used in schools and in classrooms; on the use of performance tests in high-stakes environments; on the impact of language skills on performance assessments; on the use, extent, or impact of minimum-competency tests; on the use of curriculum-embedded tests; on the factors that produce higher levels of test-score inflation; evaluating testing programs before or after implementation; or on the topic of test-score inflation prior to their 1990 study.

Despite all the outrage and scolding I was never fired. My work and productivity were outstanding. No matter how much they may have wanted to, scolders could not fabricate a legal case for firing me.

But, I was laid off. Those unemployment checks were welcome, too, as it was during periods of unemployment that I managed to find the time to complete my most important research and writing.

Certainly, I am much less wealthy than I might have been had I behaved as the scolders wanted, but I am fine.

US education research is not.

Larger Implications

I once submitted a manuscript to a journal with evidence that contradicted a claim made by the well-known psychometrician, Robert Linn, longtime co-director of CRESST. One reviewer insisted that the article be rejected out of hand. The one-paragraph review was of the “How dare you!” variety. The review asserted “no one has done more for psychometrics over the years than Robert Linn.” The reviewer did not address the substance of the manuscript.

I had not criticized Robert Linn, whom I had never met, nor had I broached the topic of his contributions to the profession. I had simply presented evidence that contradicted a single research conclusion that he had made.¹¹ But, that was enough to end consideration of my submission.

b) Standardized educational testing, particularly when it has stakes, is enormously costly in monetary terms.

c) All high-stakes testing is prone to “test-score inflation”—artificial rises in average test scores over time due to “teaching to the test.”

d) That teaching-to-the-test induced test score inflation is highly variable and sometimes massive.

e) No- or low-stakes tests, by contrast, are not susceptible to test-score inflation (because “there are no incentives to manipulate scores”).

f) Whereas score trends for high-stakes tests are unreliable, those for no- or low-stakes tests are reliable.

g) No- or low-stakes tests may be used validly as shadow tests to audit the reliability of high-stakes tests’ score trends.

h) The cause of educator cheating in test administrations is high stakes; without high-stakes, educators do not cheat.

i) The evidence is overwhelming that testing does not improve education.

¹¹ Linn's famous article "Assessments and Accountability" implies that differences in 9- and 12-month federal collections of Title I testing results around 1980 supported the notion that

At least that reviewer reported his objection honestly. Most with similar feelings contrived more superficially reasonable excuses.

If it is not possible for one to critique other research and succeed—or even remain securely employed—in a research profession, how is the profession ever to rid itself of flawed, biased, or fraudulent research? Answer: it will not.

Any community that disallows accusations of bad behavior condones bad behavior.

Any community that disallows evidence of falsehoods condones falsehoods.

In-group leaders can promote falsehoods as truth because a volunteer army of enablers protects them, stamping out any dissident embers as they appear. Most of those who recognize the falsehoods say nothing, given the rational fear of retribution and career stunting.¹²

In his book, *Skin in the game* (2018), the statistician and philosopher Nassim Nicholas Taleb makes these relevant points:

"Minorities, not majorities, run the world. The world is not run by consensus but by stubborn minorities imposing their tastes and ethics on others.

"Ethical rules aren't universal. You're part of a group larger than you, but it's still smaller than humanity in general.

"True religion is commitment, not just faith. How much you believe in something is manifested only by what you're willing to risk for it.

"Intellectual monoculture prevails in the absence of skin in the game."

Students, families, and US society as a whole pay the price for US education research's falsehoods, doggedly defended by an aggressive minority in-group, while most education researchers pay no price at all. The falsehoods persist because few of those who know they are false summon the courage to speak up. The implicit pact between the aggressive minority and the passive majority sustains an intellectual monoculture.

high-stakes cause test-score inflation. But, he neglected to account for summer learning loss, and the testing had no stakes—states and districts were required to report results, nothing more.

¹² Which, of course, leaves the few of us who have "stuck our necks out" that much more exposed and vulnerable.

Frustratingly, the mainstream education press tends to accept what the powerful elites present them at face value, as if US education research were just like that in any other subject field, such as physics or business administration. Indeed, some education reporters casually dismiss reports of censorship, thus ignoring (and supporting) the problem. They seem to assume that such widespread and successful information control could not be possible. (See, for example, Russo 2015.)

Should society at large feel OK with this arrangement? In part, Scolding #9 concerned the work of arguably the country's most policy-influential scholar in educational testing. I maintain that his primary contribution to the research literature is not just highly misleading, but fraudulent—he makes claims that he must know are false, yet he recommends basing highly consequential public policies on his work.

Yet Scolder #9 talked down his work as a "pet theory" that "he may go overboard on." In other words, not to worry, we can just hide this under the covers inside the profession's tent. Any effect on the public is no concern of ours. Education research is ours to own and manage as we please, and none of the public's business.

Yet, the society of education research professionals also wants the public to believe that they can be trusted as arbiters of truth in education policy, even though, for some, truth ranks among the least of their interests. Outing truth requires free inquiry, open debate, vigorous discussion, and conflict, all anathema to those professionals more concerned with their personal career trajectories, the superficial appearance of decorum, or the preservation of appealing myths.

References

- Haight, J. (2017, December 17). The age of outrage: What the current political climate is doing to our country and our universities. *City Journal*. New York: Manhattan Institute. <https://www.city-journal.org/html/age-outrage-15608.html>
- Kramer, R. (1991). *Ed school follies: The miseducation of America's teachers*. New York: Free Press.
- Olson, W. (2017, April 10). Shouting down speakers—a regular, organized campus business. *Minding the campus*. <https://www.mindingthecampus.org/2017/04/10/shouting-down-speakers-a-regular-organized-campus-business/>
- Phelps, R. P. (1996, March). Test-basher benefit-cost analysis, *Network News & Views*. <https://richardphelps.net/ED402330.pdf>
- Phelps, R. P. (1999, April). Education establishment bias? A look at the National Research Council's critique of test utility studies. *The Industrial-Organizational Psychologist*, 36(4) 37–49. <https://www.siop.org/TIP/backissues/Tipapr99/4Phelps.aspx>
- Phelps, R. P. (2003). *Kill the messenger: The war on standardized testing*. New Brunswick, NJ: Transaction, pages 127–146.
- Phelps, R. P. (2008/2009a). Educational achievement testing: Critiques and rebuttals. Chapter 3 in R. P. Phelps (Ed.) *Correcting Fallacies about Educational and Psychological Testing*. Washington: American Psychological Association.
- Phelps, R. P. (2008/2009b). The rocky score-line of Lake Wobegon. Appendix C in R. P. Phelps (Ed.), *Correcting fallacies about educational and psychological testing*, Washington, DC: American Psychological Association. <http://supp.apa.org/books/Correcting-Fallacies/appendix-c.pdf>
- Phelps, R. P. (2014, October). Review Essay: *Synergies for Better Learning: An International Perspective on Evaluation and Assessment* (OECD, 2013), *Assessment in Education: Principles, Policies, & Practices*. doi:10.1080/0969594X.2014.921091 <http://www.tandfonline.com/doi/full/10.1080/0969594X.2014.921091#.VTKEA2aKJz1>
- Phelps, R. P. (2015, Summer). The gauntlet: Think tanks and federally funded centers misrepresent and suppress other research: *New Educational Foundations 4*, <http://www.newfoundations.com/NEFpubs/NEFv4Announcement.html>

Phelps, R. P. (2016, March). Dismissive reviews in education policy research: A list. *Nonpartisan Education Review/Resources*.

<https://nonpartisaneducation.org/Review/Resources/DismissiveList.htm>

Russo, A. (2015, June 15). An "extremely" critical take on education trade journalism. *Phi Delta Kappan*. <https://www.kappanonline.org/an-extremely-critical-take-on-education-trade-journalism/>

Schalin, J. (2019, February). *The Politicization of University Schools of Education: The Long March through the Education Schools*. Raleigh, NC: James G. Martin Center for Academic Renewal. <https://www.jamesgmartin.center/wp-content/uploads/2019/02/The-Politicization-of-University-Schools-of-Education.pdf>

Staradamskis, P. [alias] (2008, Fall). Measuring up: What educational testing really tells us. Book review, *Educational Horizons*, 87(1).

<https://nonpartisaneducation.org/Foundation/KoretzReview.htm>

Talib, N. N. (2018). *Skin in the game: Hidden asymmetries in daily life*. New York: Random House.

Appendix: Scoldings, 2000–2009

The ten scoldings described below are eyewitness accounts from their target. Though common to one person, they may well illustrate the range and variety extant in US education research. It is highly doubtful that I am the only person in the country with such experiences.

Moreover, these examples demonstrate how deeply embedded intolerance, censorship, and information suppression are. The scolders represent a broad cross-section of respected and influential professionals.

To verify the latter claim, I list the names of the scolders in alphabetical order by last name at the end.

Scolding #1 (2000).

The year 2000 was a presidential election year and, probably, the first in which standardized testing emerged as a prominent campaign issue. The press was all over it, close to uniformly condemning standardized testing in general, and the testing program in the state of Texas in particular. (Candidate George W. Bush had been governor of Texas.) Every week seemed to release a new anti-testing book written by an activist or education school professor.

To help balance coverage of the topic, I assembled some policy-relevant and time-sensitive research. I could have published the work myself as, it so happens, I ended up doing anyway. But, I thought the work would get more traction from a sympathetic organization with a higher profile.

I sent the research to a nationally known advocacy organization to use as it saw fit, but then heard nothing from them for weeks. Meanwhile, they published other research on the same topic. I wrote to inquire what had happened to what I sent, and why they had not used it. It was an innocent question; I wanted to know if I should bother communicating with them in the future.

I received a reply from one of their research analysts. His answer had nothing to do with the research material I sent. Rather, he wrote that he had been a senior editor at a national education news publication and had inquired about me at both his current organization and among his colleagues at his former news publication. No one at either place had heard of me. Ergo, anything I sent them was not worth wasting any of their time on. It wasn't that what I had sent them that didn't matter. What didn't matter was me. I was simply not important enough to merit a moment of their attention.

Scolding #2 (2001).

I interviewed for a position with a large D.C.-area firm that specialized in employment (a.k.a., personnel, industrial-organizational) testing. The series of

meetings with potential workmates lasted most of the day. I also delivered a well-received auditorium presentation to most of the staff. The several individual and small-group interviews throughout the day went well. I enjoyed lunch with several employees. All conversations proceeded splendidly.

The firm had just won a large contract and was hiring several new staff. I judged the odds of a job offer to be very high. As friendly as could be, the president of the firm interviewed me early in the day and also introduced my presentation.

Then, later in the afternoon, in between the second-to-last and last interviews of the day, the president of the firm approached me, copy of a journal in hand. He had noticed from my résumé that I had published an article in the industry journal, *The Industrial-Organizational Psychologist* (TIP). The article critiqued a particular National Research Council (NRC) report, highly unpopular among personnel psychologists. That report had unfairly derided some of the most important and respected research in the field, which I defended (Phelps 1999).¹³

The president of the firm approached me, visibly angry with a flushed complexion, pressing his finger on the list of committee members in the NRC report. I had claimed in my article that no academic experts in the report's topic were included among the committee members responsible for the NRC report. The president was pointing to one name in the list—that of the one topical expert on the committee, who was working in industry at the time. He thought I was unacceptably wrong. But, whether one counted one or zero, the committee included few topical experts, and an oddly large number (six) of educational testing scholars. In the article, I also described some of the latter group's research to illustrate their predisposition against high-stakes testing. (See, for example, Phelps 1996.)

He defended the NRC report, and seemed to have had a part in its genesis himself.

Turns out, the firm's president, though himself a personnel testing expert who managed a consulting firm that specialized in personnel testing, also on occasion bid

¹³ My critique concluded that the NRC report was not only very poorly done, but clearly biased. Indeed, it appeared to have been deliberately arranged to deliver a negative appraisal of a huge personnel psychology research literature. Literally thousands of research studies were declared not to exist. The 14-member committee included a few psychologists, but only one who was expert in the report's topic. Among the non-experts were several scholars from the field of K-12 educational testing, prominent opponents of high-stakes and externally administered tests (Phelps 2008/2009b). A 29-member "Liaison Group" was assembled, filled with the topical experts that one reasonably would have assumed would be included as authors. But, the Liaison Group was never consulted and had zero influence on the content of the report. Yet, their existence and membership were prominently displayed in the published volume, as if to give the appearance that they participated in the production of a report that denigrated their research.

on educational testing contracts, and served on committees with the educational testing scholars of the NRC committee. I had the temerity to challenge some of the most transparently biased works in the education research literature, and that upset him.

During my last job interview of the day, an employee informed me that the president was a longtime friend of some of the education professors on the NRC panel that had unfairly panned the personnel psychology research literature. One might argue that he betrayed his primary profession for the sake of friendships or greater career opportunities. (Indeed, some years later he was elected president of the primary organization of *educational* testing professionals.)

I sent the standard thank you letters to all of the firm's employees with whom I had interviewed, but received no further communication from the firm.

Scolding #3 (2001).

I was included in a conference panel on trends in testing, put on by a test industry association. Before it was my turn to speak, another speaker—a journalist—reported on a survey her newspaper had conducted that found that the American public was opposed to the practice of using a “single standardized test” to make graduation decisions. Her organization urged policy makers to pull back on the testing, as it appeared that the public thought they had gone too far.

When it was my turn to talk, I offered that there was no state in the country that used a single standardized test for graduation decisions and probably never had been. All states allowed candidates several, many, or an unlimited number of retakes to pass. The tests were not timed. The tests were set at a 6th- or 7th -grade level of difficulty. And, there was no state in the country that required only a test for graduation: there were attendance requirements, course accumulation requirements, course distribution requirements, community service requirements, and so on. Anyone failing just one semester of English or Physical Education could not graduate in most states, no matter how well they performed on their graduation exam. All states used multiple measures—always had, and probably always would.

After the talk, the other speaker, visibly agitated, cornered me and accused me of making “an *ad hominum* attack.”

In fact, I had said nothing personal about the speaker, whom I did not know and had not before met. Afterwards, I asked my spouse and others at our lunch table if I had sounded threatening. They replied that I had spoken so softly that, sitting at the back of the room, they could barely hear me.

After the conference, the journalist reportedly confronted the people who had misled her with the “single standardized test” notion. They worked at CRESST.

Scolding #4 (2002)

While working for a research firm, I was assigned to work on an "outside," "independent," "third-party" evaluation of a large-city school district's financial condition. The district was in financial trouble and wanted more money from the state. Our clients for the evaluation were the district and the state.

Despite thousands of pages of analysis, the problem boiled down to a single issue. Like many rustbelt, formerly manufacturing dependent big cities, this one had lost population and, with it, student enrollment. Out of necessity, it had closed schools and laid off teachers and other school-level personnel. It balked, however, at laying off district-level administrators. Their ratio of district-level administrators to students loomed four times higher than the state average. They also wished to build a new school in a gentrifying neighborhood.

This administrative overhang caused problems beyond the obvious salary and benefits burden. One day, while interviewing a school principal, she received a request from one of the district's several geographically dispersed administrative offices for some information about her school—information that would require a not insubstantial amount of her staff's time to produce. Most frustratingly, however, she had received an almost identical request earlier that morning from another administrative office in another part of the city.

Our school district client, however, pressured us to declare that they were strapped for money through no fault of their own, had no remedy available, and should be granted extra subsidies from the state. At least one other contractor involved in another aspect of the evaluation had caved in and altered the language of their report to please the client.

My boss, apparently, wished to follow suit, which, for me, would have meant burying the district-level administrator overhang issue and declaring a critical need for state funding. She met me in my office, closed the door, and made her pitch. I asked her if she wanted me to write what I considered to be untrue. She said no. She wanted me to follow my professional instincts, but she thought that meant going along with her. Would she put her request in writing? No. If I wrote what she seemed to want me to write, would she be kind enough to remove my name from the report? No.

The conversation, and the relationship, slid downhill from there.

Scolding #5 (2002).

I wrote a critique of a book-length journal article that I considered not only very poorly done, but also clearly fraudulent. The author had mis-cited sources, surreptitiously altered the definitions of terms, altered some data, and made dozens of calculation errors. Moreover, all the "mistakes" led in the same direction, strongly suggesting that they were deliberate (Phelps 2003).

I had taken the time to check all of the author's claims, all his data, and all his calculations. Of the dozen or so research conclusions the author had made, I found none that stood up to any scrutiny.

I sent my critique to the editor of that journal, one of the world's most prominent mainstream education journals. I never heard back from him, at least not directly. The editor did, however, have conversations with my superiors at my place of employment during which, apparently, he suggested that I be fired.

An abrasive, accusatory scolding from the head of my division suggested that they intended to honor his wishes. The editor was a world-renowned researcher; one of education's most celebrated. And, my firm wanted him to join them on a bid for a large contract. Appeasing his prejudice was apparently more important than dealing fairly with me.

Luckily for me, I was already on my way out the door, having accepted a research fellowship elsewhere.

The fraudulent article, however, has now been cited hundreds of times as "evidence" of this or that assertion about education policy and practice. The journal never published my critique. It did, however, publish an abridged version of another scholar's critique of a single aspect of the fraudulent article, two years after the editor received it, and long after the topic in question had faded from public attention. Some appearance of open-mindedness and tolerance for divergent points of view was maintained.

Scolding #6 (2003)

The one-year fellowship at a test-development organization allowed me to pursue my own research, with access to some of the profession's brightest minds and best resources, while also receiving a generous salary. It was a wonderful experience. My office adjoined the organization's public policy and assessment validity units. My co-workers publicly praised my work ethic and productivity.

The manager responsible for the fellowship program as a whole, however, worked in a different part of the organization. One day, he called me to his office across campus. In anticipation of the meeting, I prepared a progress update on my work at the firm. As it turned out, our meeting had nothing to do with my work at the firm.

The manager had done some checking into my past, private investigator-like, and learned that I had written a report for the allegedly conservative-leaning think tank, the Thomas B. Fordham Foundation. The report was about standardized testing and highly supportive of it, countering some of the most popular criticisms arrayed against it.

The manager and I worked for one of the world's most prestigious testing firms. One might surmise that he would be grateful for the report I wrote, which validated his career and the work of his employer.

But he, personally, had had some disagreements with the Fordham Foundation and, in particular, a staffer there named Michael Petrilli, someone I had neither met nor worked with. I was not Michael Petrilli, but I would do over the course of the next hour to absorb this manager's invective about Petrilli's alleged bias, poor work and behavior.

During the emotional dressing down, the manager answered all his work and personal telephone calls, as if to emphasize how relatively unimportant I was.

My Fordham report had also disagreed with the work of the Center for Research on Educational Standards and Student Testing (CRESST), a frequent opponent of standardized testing programs, but a wealthy and influential force in the profession, with whom this manager apparently wished to be on the best of terms. The co-director of the criticized CRESST work was one Robert Linn, arguably the country's most influential testing policy scholar at the time, who was also apparently very well liked by those who knew him personally.

At the penultimate moment in my scolding, the manager inserted what he seemed to regard as irrefutable proof that I must be wrong: "I'm told that you even criticize Bob Linn." I had not personally criticized Bob Linn, but I had disagreed with some of his research conclusions.

Scolding #7 (2003)

Here is an excerpt from a review of my book *Kill the Messenger: The War on Standardized Testing* (2003), from the American Library Association, as published in their review journal *Choice*

"With an educational viewpoint shaped by socially and economically conservative ideologies, ...[the author] repeatedly asserts that economists, psychologists, and 'testing researchers' should have a voice in the conversation while teachers, school administrators, and teacher educators should not, because they have a vested interest in preserving the *status quo*. ...Summing Up: Not recommended."

To the contrary, *Kill the Messenger* repeatedly asserts that everyone should have a voice in the conversation but that, at that point, only the vested interests had one. Nowhere does the book advocate replacing one type of censorship with another. Nor have I ever advocated that any point of view be muffled; my cause, for thirty years now, has been exactly the opposite.

The most amusing criticism in the review was that I am politically conservative. Naturally, the reviewer knew nothing of my politics and nothing about how my

“educational viewpoint” had been “shaped.” I have always considered my advocacy against censorship in education to be associated with a fondness for consumer rights, the public’s right to know, transparency in the administration of public institutions, and quality control over the use of public resources. How one comes to classify those predilections as “socially and economically conservative” is unknown to me.

Scolding #8 (2004)

I submitted a manuscript on public opinion of standardized testing to one of the profession's primary scholarly journals. The editor insisted that the manuscript could only be published as a “commentary,” and not as a research article. Why? Here's one comment:

“Some word choices and phrases seem unnecessarily provocative, even a bit inflammatory. Unnamed ‘prominent educators’ are mocked a bit for ‘describing...placard-waving students and parents taking to the streets.’ Some of the writing in this paper has a tone of thumbing eyes. ... my responsibility as editor is to promote reasoned and respectful debate.”

I agree with some of what the editor writes. The language was provocative, and perhaps inflammatory. But, it wasn’t *my* language. I had quoted (but not cited) testing opponents or paraphrased what they had written in order to tone down *their* provocative, inflammatory language.

His comment, moreover, referred to just six lines in a 33-page document of empirical results. He could have simply suggested that I delete those six lines.

Classifying the article as a “commentary” would absolve him of any responsibility for its content, conclusions, and publication. But he also insisted that I make extensive, time-consuming alterations. All that before he would send the paper—the alleged “commentary”—out “for blind review.” That is, he would choose others, unknown to me, who could also insist on changes while assuming no responsibility for the result. Being a commentary published under my name, I would assume total responsibility for a product that the editor and his chosen reviewers had extensively redone.

I took a pass.

Scolding #9 (2008)

By this time, I was well aware of the ripe sensitivities to critiques of *status quo* education research. So, before accepting a position with another test development firm, I offered to cease all writing on testing policy, or to continue writing only with a pseudonym. The co-directors of the department in which I would work rejected my offer. They preferred, instead, that I keep writing and publishing outside of work, but keep them apprised.

My job candidate talk critiqued the CRESST narrative that "high stakes cause test score inflation" (i.e., artificial test score gains). My talk empirically contradicted some of their supporting arguments. I got the job.

A year-and-a-half later, I showed my co-bosses two of my soon-to-be-published pieces, one a review of a book written by a CRESST author (Staradamskis 2008)—I was the book review editor of a popular education journal at the time—and the other my chapter in a volume on testing opponents' fallacies that I was editing for the American Psychological Association (APA) (Phelps 2008/2009a). Both pieces had been well foreshadowed by my previous decade of writings and presentations and my successful job candidate talk.

Nonetheless, one of the bosses expressed outrage and insisted that I retract them both. In the case of the APA book chapter, it had been two years in the making, had emerged from several rounds of review by six different reviewers and editors, at, arguably, the world's foremost, most trustworthy source of psychology research. Moreover, I had signed a contract to produce the book.

Each piece of writing received its own separate, seemingly interminable inquisition.

Attempting to discern what was really bothering the scolder, I asked if she believed the CRESST research and narratives that I was critiquing. She replied that that was not the point. Did she, perhaps, think that critiquing CRESST's work would be bad for the firm? I doubt it; my chapter was highly supportive of our organization's work.

But, she refused to explain her motivation, hiding behind feigned outrage and an insistence that the form and tone of my writing were inappropriate. For the book review, for example, she insisted that book reviews could exist only in a certain form and format, such as that one might have learned in high school English class. I was the book review editor of one of the country's more popular education publications; she was not. But, she insisted that any book review must be structured her way, presumably even the dozens written by others for the same publication.

Reading between the scolds and the insults, her primary concern seemed to be for her own professional relationships. Though she did not work directly with the folk at CRESST, she worked regularly with people who liked both them and their work. Those relationships did not exist within our workplace. They were outside professional relationships—with colleagues on editorial boards and association committees, friends made at conference gatherings, and old graduate school buddies.¹⁴

¹⁴ The society of educational testing experts—education psychometricians—may be one of the professional world's coziest. After doctoral-level training in one of only a few dozen programs, 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

I offered to attach disclaimers to the effect that my writing was my own and did not represent the organization. I offered to use an author pseudonym. Neither was enough for her. Was there something I could change in the publications to appease her? No, she wanted them either completely neutered or wholly destroyed.¹⁵

By itself, the tête-à-tête between the angry manager and me was probably of little import. As dramatic and unreasonable as it may have been, her behavior was normal for her. The detrimental effect occurred with her co-manager. He knew little of the details or context which the two of us were debating and, for his part, had approved all my writings without equivocation. But, cueing off his co-manager's vitriol, he surmised that I must have been doing something wrong. Or, he may simply have wished to appease the more important person in his professional life by siding with her from then on.

Several months later I lost my job despite a remarkable run of productivity, producing in my last year *five times* the work that my division was originally asked to complete. As for my grand inquisitor, a year later she would receive a lifetime achievement award from the profession's primary member association.

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 organizations 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.

Indeed, I would argue that many US educational psychometricians feel so collegial that they refuse to criticize each other's work. All work conducted within the fraternity, even if wrong, biased, or fraudulent, is tolerated. To keep the profession humming along, with no threatened egos or open conflicts, all research results are accepted, or at least all those from the more powerful and well connected.

The no-criticisms-allowed culture may be understandable from a career-professional perspective, but not at all from a societal or scholarly perspective. That's because the same researchers who smother conflict within the profession, and tolerate strikingly contradictory conclusions within the profession's body of work, also seem to feel no contradiction when they serve as allegedly objective guardians of the profession's research literature and its standards for truth, research integrity, and public policy.

¹⁵ I had not originally planned to affiliate myself with my workplace in either of the publications. None of the work for the publications had been done on company time or represented the company's official point of view. But, ironically, when my inquisitors finally concluded after hours of haranguing that I would not do as they wished, they insisted that I include my company affiliation. I did so with the APA book chapter; the book review I published under a pseudonym.

Scolding #10 (2009)

The powers-that-be at the testing firm decided to hire a new director for the research division. I was asked to assist in interviewing one of the candidates. She seemed an odd choice. She had recently completed her PhD not in psychometrics, but in the philosophy of education, with a dissertation on the philosopher Wittgenstein. Her C.V. fit easily on one page. Yet, she had recently been appointed head of research for the City of Chicago Schools. A group of us spent an hour-and-a-half with her, revealing emphatically that she knew next to nothing about the work we did.

After the interview, the vice-president in charge of the hiring visited separately with each of the interviewers. As he and I were walking down the hall toward my office, I offered a few observations: psychometric research and analysis represented the core competency of the company; psychometrics was a very technical, complicated topic with which she seemed only tangentially familiar; and hiring her would mean that the two line directors overseeing our research division would both lack psychometric training or experience (as his experience and training was also in an unrelated field; he was an expert in strategic partnerships).

It was no secret that he was not a psychometrician. But, perhaps, my reminding him struck a raw nerve of insecurity. Once inside my office with the door closed, he proceeded to shout at and insult me. He also divulged the rationale for hiring the Wittgenstein scholar to run one of the world's premier psychometric research shops: apparently, she was a close friend of Arne Duncan, who had recently been appointed US Secretary of Education.

I offered that the company could hire her at a high level that did not have direct line management responsibilities, just as other companies in our industry did routinely. I would learn later that he lobbied to have me removed from the firm.

Scolder Index**Scolder (#scolding)**

Anonymous reviewer (7)
Steve Ferrara (8)
Joy Frechtling (4)
Drew Gitomer (6)
Gene Glass (5)
Craig Jerald (1)
Lynn Olson (3)
Nancy Peterson (9)
Steve Robbins (9)
Ranjit Sidhu (10)
Laurens Wise (2)