



The Gauntlet: Think Tanks and Federally Funded Centers Misrepresent and Suppress Other Education Research

by *Richard P. Phelps*

I BECAME INVOLVED IN the education-testing debate purely by chance. I did not begin as an advocate for standardized testing. And, truth be told, I am still not motivated primarily by a fondness for standardized testing, despite the fact that I have come to deeply appreciate its benefits and strengths. I am strongly motivated, however, to battle deliberate misrepresentation, censorship, and information suppression.

THE GENERAL ACCOUNTING OFFICE

More than two decades ago, while working at the U.S. General Accounting Office (GAO, now called the Government Accountability

Office), I completed a study that measured the extent and cost of standardized testing in the United States (U.S. GAO). The first President Bush, George H. W., had proposed a national assessment system that would test U.S. students in five core subject areas at three grade levels. You probably have not heard of the proposal because it died a natural death after President Bush lost his re-election bid in 1992. Part of my job at the GAO was to estimate the proposed new testing system's overlap with current testing—the time and cost it would add. In the process, I would also build a highly detailed database of state and local district assessment practices based on the GAO data collection.

We did an exceptionally thorough job. We developed surveys carefully, reviewed and pretested them, and through enormous persistence, achieved very high response rates. We collected budgets from most states and many school districts to use in benchmarking the survey results. A “Who’s Who” of notables in the evaluation, statistical, and psychometric worlds (e.g., William Kruskal, Lee Sechrest, Mark Lipsey) reviewed various aspects of the study. Nothing like it in quality or scale had ever been done before, or has been done since.

The many peer reviews from both inside and outside the GAO were rigorous, as one would expect for an investigation into a key aspect of a major presidential proposal. On all GAO quality measures (e.g., survey response rates, fact-checking) the study exceeded GAO norms.

The study results, however, were surprising, at least to me. I had been led to believe by the most accessible education-policy literature that education testing was exceptionally costly and time-consuming. The evidence showed that it wasn't, even when one accounted for all the opportunity costs in personnel time at all levels—national, state, school district, school, and classroom. In 1990–1991, system-wide (i.e., external) testing and test-related activity made up on average about seven hours per year of a student's time and about fifteen dollars in purchase costs and staff time.

The results surprised others as well. One outside review provided my first taste of a type of reaction, one more emotional than substantive, that would later become very familiar. My results could not possibly be correct, went the argument: I must have left something out. Tests cost more and take up more educator time than I had found, this reviewer was certain: additional calculations were needed, which I made, but my critic judged them unsatisfactory as well.

For those unfamiliar with such research, judgments of its quality and the trustworthiness of the results are typically benchmarked by two aspects: the size and representativeness of the sample of relevant units—public education administrative units in this case—and the scope of the measures (i.e., accounting for all relevant components of cost and time). I made every effort to ensure that not a single relevant cost or time component was neglected and conversely that no extraneous cost or time components were included.

Since then, as far as I can tell, no study of the extent or cost of testing in the United States has come anywhere close to matching the scale and coverage of the GAO study. Forty-eight states that used testing programs in 1990–1991 as well as more than six hundred school districts—a robust, nationally representative sample—had delivered complete survey responses.

Most studies undertaken since then have reported partial information: for the state level only, from a few to several school districts only, or for the purchase costs of tests and test-contractor services only (not the opportunity costs of education personnel time).¹

The GAO, however, has a single client—the U.S. Congress. Once a report has been presented to Congress, no further effort at dissemination is made.

TREATMENT OF THE GAO REPORT

Case One: The Center for Research on Educational Standards and Student Testing (CRESST)

I left the GAO before the report was actually released in January 1993; pressure to suppress the report and its findings—essentially, that standardized testing is not excessively burdensome or expensive—apparently began even before its release.² Over the ensuing months, I learned of additional efforts to suppress or misrepresent the report's findings. Conference panels, to which I was not invited to participate, criticized the report. Reports written by the federally funded Center for Research on Evaluation, Standards and Student Testing (CRESST) and elsewhere lambasted the report and suggested that better studies were needed.³ The critics claimed that the GAO report omitted information that in fact was not, and that it included information that in fact was not. But reasonable people who heard CRESST et al.'s version of the story believed it, so the GAO report, along with probably the most thorough and detailed database on

state and local testing practices ever developed, began fading into obscurity.

In place of the GAO study, other conference presentations and journal articles in mainstream education journals purported to show that standardized tests cost an enormous amount and overwhelm school schedules with their volume. Other 1990s-era studies were based on tiny samples: a single field trial in a few schools, a few telephone calls, one state, or in some cases, facts that were just invented. The cost studies among them that actually used some data for evidence tended to heap all sorts of non-test activities into the basket and label them costs of tests.

The two testing-cost studies that CRESST promoted in three successive annual conferences were based on a tiny sample (from a New Standards Project field trial) and a single state (Kentucky; Picus and Tralli). In the latter, survey responses were apparently accepted *as is* without review: for example, they included a response claiming that salaries of school personnel for the entire school year should be considered test preparation and added to the cost of tests. Both studies were widely praised and disseminated. The first of the two studies was summarized and published as the lead article in a 1995 issue of the *Journal of Education Finance* (Monk, 1995), along with misrepresentations of the GAO report.

Giving such work the benefit of the doubt, those authors may have merely misread the GAO report's specifications of the opportunity costs of personnel time. The opportunity costs of testing, however, are noted starting on page 1 and on most pages thereafter. They are noted in the introduction; the conclusion; and every chapter in between. They are included in many of the figures and tables.

I wrote dozens of letters and made dozens of telephone calls to the researchers of the testing-cost studies mentioned above; to those responsible at the organizations promoting their work; and to the U.S. Education Department (US ED), which funded (and continues to fund) CRESST. At one researcher's request I furnished him with technical documents and instruments from the GAO project work. In most cases, I was simply ignored. In a few cases, I received assurances, first, that the matter would be looked into—it was not—and second, that an *erratum* would be published in the CRESST newsletter; it never was. Articles I submitted based on the GAO study were rejected by mainstream education journals for outlandish and

picayune reasons, or because “everyone knows” that the GAO report was flawed.

The response from the relevant U.S. Education Department program officer was particularly revealing. CRESST has operated for three decades under repeatedly renewed federal grants. Consequently, no other federally funded research center has focused on testing policy. Those many millions of federal dollars have granted CRESST directors and affiliated scholars enormous power to decide which and whose research becomes known and which and whose does not. I complained to the U.S. ED grant program officer that CRESST had misrepresented the GAO report at three successive annual conferences, denied my request to attend, and ignored my requests to add errata in their publications. CRESST, I was told, was responsible for any “editorial” matters.

The trend continued even when I was finally allowed to present the results of the GAO study at an education-research conference (Phelps, 1998). During the question-and-answer session following my presentation, one individual standing at the back of the room suggested that the study’s failure to address opportunity costs deprived it of any value. I asked my questioner to identify which costs were left out, but he did not respond and soon left the room. The damage had been done—the misrepresentation of the GAO study had once again been reinforced.

Finally, I decided to send the *Journal of Education Finance* a commentary rebutting such misrepresentations as a response to a lead article the journal had published in 1995, but my initial approaches were rebuffed. I then contacted the chief editor of the journal directly. She approved the manuscript for publication and provided space for her board member to respond (Monk, 2006; Phelps, 2006). In my space in the school finance journal, I criticized the disparagement of the GAO report as censorial and its misrepresentations as tending to discredit it. The response? My criticism of the disparagement was itself censorial.

The critics continued their assault after publication of the commentary-response. Two years later my other critic from CRESST published another report, with the misrepresentations intact (Picus and Tralli). I managed to get one offending paragraph excised, but several others remained. Ultimately, I wrote an article summarizing the methods and results of the GAO report, which won two national prizes.⁴ Later, in 1999, I updated the GAO study results with data

from 1998–1999 and inflation-adjusted cost figures, detailed the combined results in an article with up-to-date estimates of the extent and cost of testing in the United States, and submitted it to the same journal whose article a few years earlier had precipitated the rebuttal-response episode recounted above. That journal published it in its back pages (Phelps, 2000).

Case Two: The National Bureau of Economic Research

My journal article was published just prior to the 2000 U.S. presidential election campaign, the first in which standardized testing was a key issue. After the new administration took office, President George W. Bush proposed a national testing program in the accountability provisions of the No Child Left Behind (NCLB) Act. The program was modeled on one he had promoted in Texas.

As a result, the current extent and cost of testing, and any possible increase due to the president's proposal, again became national issues. Studies were conducted on some aspects of the topic, for example by Ted Rebarber of Accountability Works and the Pew Center's Stateline.org. (See Accountability Works, 2004, and Danitz.)

The most widely publicized testing-cost report from the early 2000s, however, came from Carolyn Hoxby (2002), a faculty member at Harvard, then Stanford, universities and the long-time director of the education program at the National Bureau of Economic Research (NBER). Her work is the best-known on the topic because of her affiliation with organizations, such as the NBER, Harvard's Program on Education Policy and Governance, and the Brookings and Hoover Institutions, that invest a great deal of money in publicity and dissemination.

I first became interested in Hoxby's work after noticing that several reports published by NBER on education topics claimed to be the first ever to study a topic or declared that no prior research on a topic existed (Phelps, 2012a). Normally, that might not seem interesting, but in each case many previous studies had been conducted.

Hoxby's own study of testing costs doesn't refer to earlier work at all. Her work is hardly noteworthy, either. She examined budgetary expenditures for testing programs from fewer than half the U.S. states. Even had she obtained them from all states, such data are problematic: some costs induced by testing end up in other categories in accounting spreadsheets, and vice versa. Moreover, Hoxby's study

took no cognizance of local school and school district costs, which can dwarf state costs.

Case Three: The National Research Council

CRESST re-entered the testing-cost debate with a report from the Board on Testing and Assessment (BOTA) at the National Research Council (NRC), a group that CRESST captured in the late 1980s and has held as its own since (Phelps 2008/2009, 2012b). The 2008 BOTA-NRC report, *Common Standards for K-12 Education?*, asserts, again, that the GAO report left something out and so underestimated the cost of testing (Beatty).⁵ And again, the assertion is false. This time, the NRC accused the GAO study of neglecting to consider the cost of “standard setting” during test development; in fact, this cost was fully accrued in the GAO calculations.⁶

Claiming a void in others’ calculations can be used as an excuse to bulk up testing critics’ own cost estimates massively. Here are just a few ways that the NRC report, *Common Standards for K-12 Education?*, overestimates the cost of testing:

- One-time-only start-up costs—e.g., standard (passing-score) setting—are counted as annual recurring costs.
- Educator travel and lodging expenses for serving on standard-setting and other test-development panels are counted twice, both as direct educator expenses and in the budget of the state education agency (which, in fact, reimburses the educators for these expenses).
- The full duration of all testing activities at a school—said to be 3–5 days—is allotted to each and every educator participating. For example, take the time of a fifth-grade teacher who administers a one-hour math exam on Tuesday of testing week and who otherwise teaches regular class that week. That time is counted as if s/he were involved in administering each and every exam in every subject area and at every grade level throughout the entire 3–5 days. Moreover, the time of each teacher in the school is counted as if the teacher is present in each testing room for all subject areas and grade levels. By this method, the NRC overestimates the educator time spent directly administering tests about twentyfold.
- Another way of looking at the problem is to ignore the fact that a school administers a series of one-hour tests across the

tested subject areas and grade levels over the span of 3–5 days and instead assume that all classes in all subject areas and grade levels spend 3–5 days doing nothing but take day-long exams—which, in fact, is not what happens.

- The NRC calculates the number of teachers involved by using a federally estimated average pupil-teacher ratio, rather than an average-class-size estimate. Pupil-teacher ratios underestimate class sizes because they include the time of teachers when they are not teaching. By this method, the NRC overestimates the number of teachers involved in directly administering tests by another 50 percent.
- The NRC counts all teachers in a school, even though only those in certain grade levels and subject areas are involved in testing—usually amounting to fewer than half a school's teachers. By this method, the NRC overestimates the number of teachers involved in directly administering tests by another 50 percent or more.
- In calculating "data administration costs" of processing test data in school districts and states, the NRC classifies all who work in those offices as "management, business, and financial" professionals who earn \$90,000 per year. Anyone who has worked in a state or local government data-processing department realizes that this classification grossly overestimates the real wages of the majority of employees, who essentially work as clerical and often contingent staff.
- The NRC is told by one school district that its average teacher spends twenty hours every year in professional development related to assessment and accountability. Despite how preposterous the number should sound, the NRC has used that one piece of hearsay to estimate the amount of time that teachers everywhere, whether involved in testing or not, annually spend in related professional development.
- Moreover, professional development related to testing and accountability is assumed to be unrelated to regular instruction, so it is counted as a completely separate, added-on (i.e., marginal) cost.
- The NRC counts educator time working on standard-setting and other test-development panels as "two or three days," which anyone who has worked in test development knows is a high estimate. One to two days is more realistic.

Finally, the NRC studied testing and accountability in only several school districts, in only three states. Nonetheless, according to the NRC, the GAO report—which as we have seen analyzed more detail from all forty-eight states with testing programs and more than six hundred school districts—is the study that left stuff out. In the end, the NRC's estimates for testing and accountability costs are, in the council's own words, "about six times higher" than previous estimates.

For several years afterward, each of the two most recognizable sides in U.S. education policy debates had its own testing costs research champion. Education reformers, think tankers, and Republican Party advocates had Carolyn Hoxby's numbers, which hugely underestimate the cost of testing programs. The education schools, educator professional associations, and Democratic Party advocates had the CRESST-NRC numbers, which greatly exaggerate the cost of testing programs. Anything in between was either ignored or misrepresented.⁷

Case Four: The Brookings Institution

These days, the education policy topic *du jour* is the Common Core Standards, and standardized testing is a key component of the planned program. Naturally, one would expect a think tank to weigh in on the matter of the possible costs, and the Brookings Institution has done so with the work of yet another Harvard University Ph.D. in economics or political science—in this case Matt Chingos, a political scientist.

Several months ago, Brookings began promoting its own report, which begins by clearing the field.

Unfortunately, there is little comprehensive up-to-date information on the costs of assessment systems currently in place throughout the country. This report seeks to fill this void by providing the most current, comprehensive evidence on state-level cost of assessment systems, based on new data gathered from state contracts with testing vendors. (Chingos, p. 1)

[Other] Estimates of these costs are based primarily on assumptions and guesswork. . . . The most comprehensive nationwide data were collected about a decade ago, in separate investigations by Caroline Hoxby and the Pew Center for the States. (p. 4)

The latter criticism—estimates “based primarily on assumptions and guesswork”—was directed at two other studies that Chingos presumably also considers not as “comprehensive” as his, cited in the accompanying footnote. A detailed reading of the Brookings report, however, reveals its own abundance of assumptions and guesswork.

Like Chingos’s own work, the Hoxby and Pew Center studies he cites examined only the direct costs of testing at the state level, not the more consequential data at the local level or any data at all on personnel time (outside the easiest-to-locate line items in state budgets). Because Chingos’s study did not examine those cost components—an absolutely necessary step for a complete cost estimate—perhaps he did not wish to draw attention to other studies that included them (e.g., *Accountability Works*, 2004; and Phelps, 2000).

As for those other cost components, Chingos pleads that they are too difficult to measure. Take for example the time spent by state employees in “selecting contractors and overseeing the vendors”:

But such costs are difficult to track consistently across states, and usually represent a small fraction of the testing budget. (p. 7)

That may fairly be termed disingenuous. State employees typically do far more than just “oversee” the vendors, and such costs are not “small,” though they may be a small fraction of the *testing budget*. The costs are absorbed in other parts of the budget—in the regular salaries for staff positions that probably would not exist if there were no testing program. Collectively, they can represent a large portion of the cost of a testing program.

The roles played by school and district officials who aid in test administration and scoring are important as well, but the cost of this work is challenging to measure. Calculating such costs requires information on which employees have these responsibilities, their compensation levels, how much time they devote to test-related activities, . . . (p. 7)

Yes, it is challenging to measure. Yes, it does require information on responsibilities, compensation levels, and time devoted to test-related activities. So did the Brookings Institution meet those challenges and gather that difficult-to-gather information? (Note: the GAO study did both.) No, the Brookings report claimed that it was too hard.

Brookings dismisses the BOTANRC cost estimates of 2008 as irrelevant because “these costs are data collected from only three states and reflect the costs of standards and accountability systems in addition to the assessment costs” (Chingos, p. 27, note 10). In fact, however, the BOTANRC estimates did not reflect the costs of standards and accountability systems in addition to the assessment costs. Those estimates had simply double counted the cost of “standard setting” (i.e., “passing score” setting) sessions. Like the National Research Council report, the Brookings report ignores how tests are actually developed.

Other excuses for not being comprehensive, even while repeatedly boasting about being *the most* comprehensive:

Time spent preparing for end-of-year tests may also be considered a “cost,” but it is one that is nearly impossible to measure given the difficulty of separating instructional time that is geared specifically towards preparation for the test as compared to for some other purpose. (p. 38, note 36)

For these contracts, we either ignore the development costs (instead focusing on the contract costs during operational test years) or divide the development costs equally over the operational years. (p. 8)

The Brookings estimates of testing costs are suspect because they are far from comprehensive. They do not include, or even attempt to include, personnel costs at either the state or the local levels. Neither do they include any local costs. Ironically, for a work that repeatedly touts its comprehensiveness, the report’s single greatest lack is comprehensiveness. (For an interesting contrast, see Accountability Works, 2012, or Nelson.)

After the truncated, skewed testing-cost estimates, all that is left of value in the Brookings report is the revelation about saving money on testing through state consortia, an idea that could have been lifted right out of the GAO report.

CRONY RESEARCH DISSEMINATION

The GAO project work was not just unfairly slighted by education’s vested interests: it was repudiated. All that effort, all that expense—funded by U.S. taxpayers—was so thoroughly and effectively discredited by its opponents that barely a trace remains in the

collective working memory of education policymakers, or anywhere else outside my own cranium and computer hard drive.

To discredit my GAO report, education's vested interests falsely accused my work of ignoring the costs of personnel time. Ironically, the think tankers' own work has comprehensively ignored the opportunity costs of personnel time and has apparently felt no obligation to include it, yet still claim comprehensiveness.

It would seem that even substandard education research from the think tanks or federally funded centers is deemed praiseworthy, while the highest-quality work from those of the vast research working classes is flicked away like a stinkbug.

This latest report from the Brookings Institution continues a twenty-first-century tradition of information suppression, misinformation, and self-promotion in education policy research from our country's best-known and best-funded think tanks. But censorship isn't the only problem: the process fosters a nonmonetary form of corruption. The currency of scholars is attention, providing the "richest" among them a confluence of honors, awards, and remuneration streams.

Both the NRC and the think tank reports mentioned above may be used to proselytize and mislead. More emphatically, they are expropriated to showcase the careers of those involved: their authors declare the work of other researchers inferior or nonexistent, while at the same time they liberally cite their own work and that of like-minded colleagues and package the combination as if it were all that mattered. The stated mandates of these organizations are to objectively review all the research available; instead they promote their own work and declare most of the rest nonexistent. They are mandated to serve the public interest; instead they serve their own.

As a result, journalists assume that the easily accessible work of think tanks and federally funded centers represents the research literature as a whole and that the dissemination of education research is objective and fair. They couldn't be more wrong.

Some journalists step further into an ethical abyss—they help promote dismissive reviews. No journalist has the time to validate such claims; it can take years to learn a research literature. When journalists mention a "paucity of research on this topic" or the like, they are probably taking one quite self-interested person's word for it. When they write "[So-and-so's] study is the first of its kind" without

further investigation, they are complicit in the corruption. Analysis and debate on education are adversely impacted at all levels—local, state, and federal.

The National Research Council's BOTA was captured decades ago by CRESST-affiliated researchers. A clique of faculty members from a handful of elite universities has monopolized the education-policy function at the country's most prominent think tanks. (Similarly, many argue that the education research function at the National Science Foundation has been captured by radical constructivists who fund the type of research they like and pretend the rest of the research literature does not exist.)

The disastrous results illustrate how federal and foundation money can concentrate power to achieve results exactly the opposite from those intended. Once small, cohesive groups control the larger organizations, they can focus their efforts on restricting entry into policy arenas to those in their own circles. The careers of those inside these groups have flourished. Meanwhile, the amount of objective information available to policymakers and the public—our collective working memory—has shrunk.

Another ramification is that too few people acquire too much influence over those who control the purse strings of education research. And those who control the purse strings wield excessive influence over policy decisions. Until the folks at the Bill and Melinda Gates Foundation and the U.S. Education Department—to mention just a couple of consistent funders of education-policy debacles—broaden their networks, expand their reading lists, and open their minds to more intellectual diversity, they will continue to produce education policy failure.

The problems of American schools can hardly be ameliorated by ignoring sound, relevant information. It would help if funds were available to a wider pool of legitimate education researchers, evidence, and information. In recent years, grantors have instead encouraged the converse—funding a saturating dissemination of a narrow pool of information—thereby contributing to U.S. education policy's number-one problem: pervasive misinformation.

SO WHAT?

Not only are these badly behaved researchers the only sources that most journalists and policymakers consult, but the effects of their bad behavior are also spreading overseas. The education-testing

research function at the World Bank, for example, has been handed down over the past few decades from one scholar affiliated with Boston College's School of Education to another. True to form, they cite the research they like, some of which is their own, most of the rest from CRESST, and imply that the vast majority of relevant research is nonexistent.⁸

Recently, the Organisation for Economic Co-operation and Development (OECD) published a study on educational assessment that followed the template of ignoring most relevant research literature and highlighting work conducted at a certain U.S. federal research center and several U.S. think tanks (Phelps 2013, 2014).

Their skewed recommendations are now the world's.

Recommended Citation: Phelps, R. P. (2015). The gauntlet: Think tanks and federally funded centers misrepresent and suppress other research: *New Educational Foundations* 4, <http://www.newfoundations.com/NEFpubs/NEFv4Announcement.html>

References

- Accountability Works, (2004, January). NCLB under a microscope: A cost analysis of the fiscal impact of the No Child Left Behind Act of 2001 on states and local education agencies
- Accountability Works, (2012, February). National Cost of Aligning States and Localities to the Common Core Standards, Boston, Mass.: Pioneer Institute.
- Beatty, A. (2008). Common Standards for K–12 Education?: Considering the Evidence: Summary of a Workshop Series. Committee on State Standards in Education, Washington, D.C.: National Research Council.
- Chingos, M. (2012, November). Strength in numbers: State spending on K–12 assessment systems. Washington, D.C.: Brookings Institution. Retrieved March 12, 2014, from http://www.brookings.edu/~media/research/files/reports/2012/11/29%20cost%20of%20assessment%20chingos/11_assessment_chingos_final.pdf
- Clarke, M. [moderator] (2013). What does the research tell us about how to assess learning? Panel discussion for World Bank Symposium: Assessment for Global Learning, November 7–8, 2013, Washington, D.C.
- Danitz, T. (2001, February 27). Special report: States pay \$400 million for tests in 2001. Stateline.org. Pew Center for the States.

- Harris, D. N., & Taylor, L. L. (2008, March 10). The Resource Costs of Standards, Assessments, and Accountability: A Final Report to the National Research Council.
- Hoxby, C. M. (2002). The cost of accountability, in W. M Evers and H.J. Walberg (Eds.), *School Accountability*, Stanford, Calif.: Hoover Institution Press.
- Koretz, D. (2013, November 7). Learning from research on test based accountability? Paper presented at World Bank Symposium: Assessment for Global Learning, November 7-8, 2013, Washington, D.C.
- Monk, D. H. (1995, Spring). The costs of pupil performance assessment: A summary report, *Journal of Education Finance*, 20(4), 363-371.
- Monk, D. H. (1996, Spring). The importance of balance in the study of educational costs, *Journal of Education Finance*, 21(4), 590-591.
- Nelson, H. (2013, July). Testing More, Teaching Less: What America's Obsession with Student Testing Costs in Money and Lost Instructional Time, Washington, D.C.: American Federation of Teachers.
- Phelps, R. P. (1996, Spring). Mis-conceptualizing the costs of large-scale assessment, *Journal of Education Finance*, 21(4), 581-589.
- Phelps, R. P. (1998). Benefit-cost analysis of systemwide student testing, Paper presented at the annual meeting of the American Education Finance Association, Mobile, Ala.
- Phelps, R. P. (2000, Winter). Estimating the cost of systemwide student testing in the United States. *Journal of Education Finance*, 25(3), 343-380.
- Phelps, R. P. (2005, September). A review of Greene (2002), High school graduation rates in the United States, *Practical Assessment, Research, and Evaluation*, 10(15). <http://pareonline.net/pdf/v10n15.pdf>
- Phelps, R. P. (2008/2009). The National Research Council's Testing Expertise, Appendix D in R. P. Phelps (Ed.), *Correcting fallacies about educational and psychological testing*, Washington, D.C.: American Psychological Association. Available at: <http://supp.apa.org/books/Correcting-Fallacies/appendix-d.pdf>
- Phelps, R. P. (2012a). Dismissive reviews: Academe's Memory Hole. *Academic Questions*, Summer. http://www.nas.org/articles/dismissive_reviews_academes_memory_hole
- Phelps, R. P. (2012b). The rot festers: Another National Research Council report on testing. *New Educational Foundations*, 1, 30-52. <http://www.newfoundations.com/NEFpubs/NEFv1n1.pdf>
- Phelps, R. P. (2013). The rot spreads worldwide: The OECD: Taken in and taking sides. *New Educational Foundations*, 2. <http://www.newfoundations.com/NEFpubs/NEFv20f0513.pdf>

- Phelps, R. P. (2014). A review of Synergies for Better Learning: An International Perspective on Evaluation and Assessment. *Assessment in Education: Principles, Policy, and Practice*, 21(4), 481–493.
- Picus, L. O., and Tralli, A. (1998, February). Alternative assessment programs: What are the true costs? CSE Technical Report 441, Los Angeles: CRESST.
- Shepard, L. (2013, November 7). How can classroom assessment inform learning? Keynote Presentation presented at World Bank Symposium: Assessment for Global Learning, November 7–8, 2013, Washington, D.C.
- U.S. GAO. (1993, January). Student testing: Current extent and expenditures, with cost estimates for a national examination. GAO/PEMD-93-8. Washington, D.C.: U.S. General Accounting Office.

Notes

1. Some have argued that an opportunity cost of student time “lost” to testing should also be included. That assumes, however, that students learn nothing when taking a test and that they would be learning something if the time were not used for testing. As it turns out, a massive research literature affirms that students are more likely to learn when taking a test (see, for example, Phelps, 2012). Hence, if the opportunity cost of student time in testing were to be considered for inclusion, it should be subtracted from the cost calculations.
2. For reasons never explained to me, the working title that I gave the study, and that had passed through all internal and external reviews—“Student Testing: Current Extent and Cost, with Estimates for National Examination”—was changed to “Current Extent and Expenditures.” This, despite the fact that we used line-item budget data—expenditure data—only to validate the survey data from state and local testing directors, which could be quite different. Line-item expenditures may or may not categorize relevant expenditures neatly; usually they do not. As it turned out, this change substantially aided the censorial efforts the leading critiques of the GAO report, which claimed that it ignored the opportunity costs of personnel time. In fact, the majority of costs in the GAO calculations were of personnel time.
3. For example: 1993 CRESST Conference (Assessment Questions: Equity Answers: What Will Performance Assessment Cost?), Monday, September 13; 1994, CRESST Conference (Getting Assessment Right: Practical and Cost Issues in Implementing Performance Assessment), Tuesday, September 13; 1995, CRESST Conference (Assessment at the Crossroads: What are the Costs of Performance Assessment?), Tuesday, September 12. CRESST report #441 still contains mostly erroneous

claims related to the GAO report, on pages 5 and 64–66, and mostly erroneous claims about CRESST’s work on the issue, in the first seventeen pages.

4. The Doctoral Scholar Award of the National Center for Education Statistics (NCES) and the New Scholar Award of the Association for Education Finance and Policy (AEFP), both in 1997.
5. On pages 8–9 of the background paper “The Resource Costs of Standards, Assessments, and Accountability” (Harris and Taylor, 2008) one reads, “On the other hand, neither Phelps nor the GAO study ascribes any costs to standard setting. . . .”
6. Test developers often confusingly use the phrase “standard setting” to identify two entirely different phases of test development. There is the writing of academic content standards and expected performance levels that takes place before the development of a standardized test even starts. Then, much later in the test-development process, after some test forms have already been administered, groups of educators, experts, and public officials gather to decide how to score the new test. Often, but not always, the “standard” being set at these meetings is the passing score for the new test, and the meetings are sometimes called “passing-score setting” meetings. But the traditional, albeit confusing, label of “standard setting” is still widely used. The GAO study included all costs for the latter type of standard setting—passing score setting—contrary to the claims in the NRC report.
7. This is hardly the only issue where education establishment and think tankers present opposing assertions as facts, with both being wrong, misleading, or exaggerated. Until the mid-2000s, for example, education establishment folk favored the use of a “graduation rate” that grossly overestimated the actual proportion of students who begin high school and later graduate. Since then, think tankers have managed to institute a different measure that grossly underestimates that proportion (e.g., by counting those who take more than four years to graduate or transfer schools as dropouts). (See Phelps, 2005.)
8. See Clarke (2013), Koretz (2013), and Shepard (2013). Long a junior partner in CRESST’s censorial efforts, the even more radically constructivist and (anti-) reliable, high-stakes testing-policy group at Boston College has somehow maintained control of the educational testing function at the World Bank for decades (viz. various works of Kelleghan, Greaney, and Clarke). Leadership succession in this office of the World Bank is not meritocratic; it is filial.