

**The Gauntlet:
Think tanks and federally funded centers misrepresent and
suppress other education research**

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, against deliberate misrepresentation, censorship, and information suppression

The General Accounting Office

Over 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 (US GAO). The first president Bush, George H. W. Bush, had proposed a national assessment system that would test US students in five core subject areas at three grade levels. You probably have not known about it because the proposal 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 a remarkably good job with that study. 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 reviewed various aspects of the study (e.g., William Kruskal, Lee Sechrest, Mark

Lipsey). 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, just 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. 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, systemwide (i.e., external) testing and test-related activity comprised on average about seven hours per year of a student's time and about \$15 in purchase costs and staff time.

Others were surprised by the results as well. One of the outside reviewers, Margaret "Peg" Goertz, from the University of Pennsylvania's Graduate School of Education provided my first taste of a type of reaction that would later become very familiar, one more emotional than substantive. My results could not possibly be correct, she argued, I must have left something out. Tests cost more and take up more educator time than I had found, she was certain. She insisted on some additional calculations, which I made, that still did not satisfy. But, she would have other, more public opportunities in the future to cast doubts on my work.

For those not familiar with research of this type, 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., were all relevant components of cost and time accounted for?). I made every effort to make certain that not a single relevant cost or time component was neglected and, conversely, that no extraneous cost or time components were included.

Since then, I have come to better appreciate that effort for, as far as I can tell, no study of the extent or cost of testing in the U.S. since has come anywhere close to matching its scale and coverage. With prodding and many follow up letters and calls, I received complete survey responses from all 48 states with testing programs in 1990–

1991, and from over 600 school districts—encompassing a robust nationally representative sample.

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

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

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

I left the GAO for other employment before the report was actually released in January 1993 and, apparently, pressure to suppress the report and its findings (essentially, that standardized testing is not burdensome and does not cost that much) descended even before it was released.² Over the ensuing months, I became gradually aware of more efforts to suppress or misrepresent the report's findings. Panels were held at conferences criticizing the report—panels to which I was not invited. Reports were written by the federally funded research center, the Center for Research on Evaluation, Standards and Student Testing (CRESST), and elsewhere, lambasting it and suggesting that

¹ Some have argued that an opportunity cost of student time “lost” to testing should also be included. But, that assumes that students do not learn anything when taking a test and they would be learning something if the time were not used for testing. As it turns out, a very large research literature affirms that students are *more likely* to learn when taking a test (see, for example, Phelps, 2012). So, if it were to be considered for inclusion, the opportunity cost of student time in testing 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 Cost 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 GAO report's chief misrepresenters—David Monk and Larry Picus—who claimed that it ignored the opportunity costs of personnel time. In fact, the *majority* of costs in the GAO calculations were of personnel time.

better studies were needed.³ The characterizations of the GAO report were completely false—the critics claimed that information was left out that, in fact, was not, and that information was included that, in fact, was not. But, reasonable people, allowed hearing only one version of the story, believed it, and the GAO report, along with probably the most thorough and detailed data base on state and local testing practices ever developed, started fading into obscurity.

After I presented the results of the GAO study at a different education research conference in 1998 (Phelps, 1998), a man standing at the back of the room suggested that the study was worthless since I had not considered opportunity costs. I asked him to identify which costs were left out, but he did not respond and shortly thereafter left the room. The damage had been done—he had suggested to many in the room that my study was critically lacking and not worth reading. Who was he? Peg Goertz's husband.

In place of the GAO study, other reports were written and presented at conferences, and articles published in mainstream education journals, purporting to show that standardized tests cost an enormous amount and were overwhelming school schedules in 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, the facts were just made up. 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 call 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). (See Monk, 1995; Picus & 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.

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.

I wrote dozens of polite letters and made dozens of polite telephone calls to the researchers of those two studies who asserted the erroneous claims about the GAO study—David Monk and Larry Picus—to those responsible at the organizations promoting their work, and to the US Education Department, which funded (and still continues to fund) CRESST. 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. I submitted articles based on the GAO study to mainstream education journals and they were rejected for outlandish and picayune reasons, or because "everyone knows" that the GAO report was flawed.

The response from the US Education Department (US ED) program officer was particularly revealing. CRESST has operated for three decades under repeatedly renewed federal grants. Consequently, it has been the only federally funded research center focused on testing policy, ever. These many millions have bestowed on CRESST directors and affiliated scholars enormous power to decide which and whose research becomes known and which and whose does not. It has also served grandly to advance the careers of CRESST-affiliated scholars.

I complained to the relevant US ED grant program officer that CRESST had misrepresented the GAO report in three successive annual conferences, denied my request to attend, and ignored my requests to add errata in their publications. He refused to do anything. CRESST was responsible for any "editorial" matters, he said, and he had no authority to intervene. This fellow has just recently retired after a few decades at US ED where he did, apparently, not much.

Ultimately, after years of being polite and following designated communication channels to no effect, I felt forced to take the issue public and wrote a commentary concerning the misrepresentation of the GAO report in an education finance journal that had published a lead article by David Monk mis-characterizing and dismissing it (Phelps, 1996). Journal editorial staff working at the sponsoring organization—The Association of School Business Officials—refused to accommodate my manuscript in any way. Perhaps that was because I, someone unknown to them, was criticizing the behavior of two education finance professors well known to them and serving on their editorial board.

Forced to end run regular channels again, I contacted the editor of the journal directly. She allowed the manuscript publication, and

provided space for David Monk to respond (Monk, 2006; Phelps, 2006). I had made every reasonable effort to inform Monk that he was misrepresenting the GAO report, including sending him technical documents and instruments from the GAO project work at his request. Still, three years passed, he continued his misrepresentation, and exaggerated the alleged uniqueness of his own and his friend Larry Picus' work.

Anyone prone to give Monk and Picus the benefit of the doubt and assume that they misread a GAO report that did not state clearly enough what components were included in the cost calculations should reconsider. It is simply not possible to read the GAO report in any depth and not understand that the opportunity costs of personnel time were included in the calculations. The fact is noted starting on page 1 and on most pages thereafter. The fact is noted in the introduction, the conclusion, and every chapter in between. The fact is included in many of the figures and tables.

Either Monk and Picus deliberately misrepresented the GAO report, perhaps in an effort to promote their own work, or they never read any of the report, despite suggesting intimate familiarity.

In my space in the school finance journal, I criticized Monk for behavior that was clearly censorial. He had blatantly and repeatedly misrepresented the GAO report, in a way that discredited it and encouraged the public to dismiss it. With Monk's response, I would receive my first taste of another kind of reaction that would later become very familiar—he accused me of being censorial simply by criticizing his work and behavior. After years' of continuous effort discrediting my work, he wrote magnanimously that everyone's work should be considered and respected.

Not even the journal commentary-response with Monk stopped their misrepresentation of the GAO report though. Larry Picus published a CRESST report two years later with all of their misrepresentations intact (Picus & Tralli). I managed to convince a new director at CRESST to excise one offending paragraph, but several others remain.

Ultimately, a paper based on the GAO report won a national prize. Later, I updated the GAO study results with data from 1998–1999 and inflation-adjusted cost figures. An article based on those up-to-date estimates of the extent and cost of testing in the United States was published in the back pages of the *Journal of Education Finance*.⁴ The

article on testing costs by Monk based on data from one ill-fated project's test field trial had been published a few years earlier as lead in the same journal.

My journal article was published in 2000, just prior to the first US presidential election campaign with standardized testing a key issue for debate. The eventually victorious Republican Party candidate, George W. Bush, proposed a national testing program modeled on one in Texas—in the accountability provisions of the No Child Left Behind (NCLB) Act.

Thus, 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 & Danitz)

The National Bureau of Economic Research

The most widely publicized report on testing costs from the early 2000s, however, was that of Carolyn Hoxby (2002), of Harvard, then Stanford University and long-time director of the education program at the National Bureau of Economic Research (NBER). Her work on the topic is the most widely known because she is affiliated with organizations that invest a great deal of money in publicity and dissemination.

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

In her own study of testing costs, Hoxby doesn't refer to prior work at all. But, her work is hardly noteworthy, either. She looked at budgetary expenditures for testing programs from less than half the US states. Even had she obtained it from all states, such data are problematic because some costs induced by testing end up in other categories in accounting spreadsheets, and vice versa. Moreover, she didn't look at all at local school and school district costs, which sometimes dwarf state costs.

The National Research Council

The CRESST folk 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 they captured in the late 1980s and have held as their own ever 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).⁴ 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.⁵

Claiming a void in others’ calculations is used as an excuse to bulk up their own cost estimates massively. Here are just several 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 (i.e., 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).

⁴ On pp. 8–9 of the background paper “The Resource Costs of Standards, Assessments, and Accountability” (Harris & Taylor, 2008) one reads “On the other hand, neither Phelps nor the GAO study ascribes any costs to standard setting....”

⁵ Test developers often confusingly use the phrase “standard setting” to identify two entirely different phases of test development. There is the development, or “setting”, 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.

The full duration of all testing activities at a school—said to be 3–5 days—is allotted to each and every educator participating. So, for example, 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, 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 and every teacher in the school is counted as if each and every teacher is present in each and every testing room for all subject areas and grade levels. By this method, the NRC overestimates the amount of educator time spent directly administering tests about twenty-fold.

Another way of looking at it 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 but, instead, assume that all classes in all subject areas and grade levels are sitting for 3–5 days doing nothing but taking 3–5-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%.

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% or more.

In calculating “data administration costs” of processing test data in school districts and states, the NRC classifies all who work in these offices as “management, business, and financial” professionals who make \$90,000/year. Anyone who has worked in state and local government data processing departments knows that this would grossly overestimate the real wages of the majority of these employees who, essentially, work as clerical and, oftentimes, contingent staff.

The NRC is told by one school district that their average teacher spends 20 hours every year in professional development related to

assessment and accountability. Despite how preposterous this number should sound, this one piece of hearsay is used by the NRC to estimate the amount of time all teachers everywhere, whether involved in testing or not, spend annually in related professional development.

Moreover, professional development related to testing and accountability is assumed to be unrelated to regular instruction and, so, 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, as 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. But, according to them, the GAO report, which analyzed more detail from all 48 states with testing programs and over 500 school districts ...is the study that left stuff out. In the end, the NRC estimates for testing and accountability costs, are in their own words “about six times higher” than previous estimates.

In addition to the usual suspects from CRESST, Peg Goertz played a key role in the presentation of the 2008 NRC report. And, recall that she was a reviewer for the original GAO report so, presumably, she had read it in detail.

For several years after, each of the two most recognizable sides in US education policy debates had their own testing costs research champion. The 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 grossly overestimate the cost of testing programs. Anything in between was either ignored or misrepresented.⁶

⁶ This is hardly the only issue where education establishment and think tankers present opposing assertions as facts, with both being wrong, misleading, or exaggerated. Up 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

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 could expect a think tank to weigh in on the matter of their possible costs, and the Brookings Institution has done so with the work of yet another Harvard University economics or political science PhD (political science in this case).

Several months ago, the Brookings Institution began promoting a report written by Matthew M. Chingos, like so many other think tank residents, a former graduate student of Paul Peterson's. Chingos begins by clearing the field before him first.

"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 snipe—"Estimates of these costs are based primarily on assumptions and guesswork..."—was directed at two other studies, that he presumably also considers to be not as "comprehensive" as his, cited in the accompanying footnote. Read the Brookings report in detail, however, and one will discover their own abundance of assumptions and guesswork.

Like Chingos, Hoxby and the Pew Center looked only at the direct costs of testing at the state-level, and not at 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). As Chingos wasn't looking at those cost components—absolutely necessary for a

that proportion (e.g., by counting those who take more than four years to graduate or transfer schools as dropouts). (See Phelps, 2005.)

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)

This is disingenuous. State employees typically do far more than just “oversee” the vendors. And these 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 Chingos and the Brookings Institution accept those challenges and gather that difficult-to-gather information? (Note: the GAO study did both.) No, they claimed that it was too hard.

Chingos and Brookings dismiss the 2008 BOTANRC cost estimates 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, footnote 10). In fact, the BOTANRC estimates did not reflect the costs of standards and accountability systems in addition to the assessment costs. They simply double counted the cost of “standard setting” (i.e., “passing score” setting) sessions. Like the National Research Council report authors, Chingos and the Brookings Institution do not seem to know the first thing about how tests are 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, footnote 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 local levels. Neither do they include any local costs. Ironically, for a report that repeatedly boasts of being the most comprehensive, the report’s single greatest lack is comprehensiveness. (For an interesting contrast, see Accountability Works, 2012, or Nelson)

After the measly and 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 they may have lifted right out of the GAO report.

Crony research dissemination

The GAO project work was not just unfairly criticized by education’s vested interests; it was annihilated. All that enormous effort, all that considerable expense—funded by US taxpayers—was so thoroughly and effectively discredited by CRESST, Monk, Picus, the Goertz’s, and other sympathizers that barely a trace of it remains in the collective working memory of education policy, or anywhere else outside my own cranium and computer hard drive.

Education’s vested interests employed the false accusation that my GAO work ignored the costs of personnel time to discredit it. Ironically, in their work, the think tankers have ignored the opportunity costs of personnel time absolutely, seem to have never felt any obligation to include it, yet still claim comprehensiveness.

It would seem that if one has been accepted into the education research aristocracy of think tanks or federally funded centers, even skimpy, shoddy work will be called great. Meanwhile, 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 faithfully continues a 21st 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 is corrupt. This particular type of corruption does not only involve money. The currency of scholars is attention, with the "richest" among them achieving the most—genuine fame—celebrity status that floods a confluence of honors, awards, and remuneration streams.

Both the NRC and think tank reports mentioned above may be used to proselytize and mislead. But, more emphatically, they are expropriated to showcase the careers of those involved. At the same time the report authors declare the work of other researchers inferior or nonexistent, they liberally cite their own work and that of their close friends, and package the combination as if it were all that matters.

Journalists, unfortunately, simply assume that the easy-to-obtain work of think tanks and federally funded centers validly represents the research literature as a whole. They simply assume that education research dissemination is objective and fair. They couldn't be more wrong.

But, 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. So, every time a journalist writes "there is a paucity of research on this topic", or the like, they're just taking one very self-interested person's word for it. Every time a journalist writes "there is little research in this area" or "so-and-so's study is the first of its kind" they are complicit in the corruption.

The National Research Council's BOTA was captured decades ago by CRESST-affiliated researchers. A small clique of faculty from Harvard and Stanford Universities has captured 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 channel money to friends and others like-minded.)

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

The stated mandates of these organizations are to objectively review all the research available; instead they promote their own and declare most of the rest nonexistent. They are mandated to serve the public interest; instead they serve their own.

Currently, too few people have too much influence over those who control the education research purse strings. And, those who control the purse strings have too much influence over policy decisions. Until folk at the Bill and Melinda Gates Foundation and the US 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.

It would help if they would fund a wider pool of education researchers, evidence, and information. In recent years, they have, instead, encouraged the converse—funding a saturating dissemination of a narrow pool of information—thereby contributing to US education policy's number 1 problem: pervasive misinformation.

So what?

The aggressive, career-strategic behavior of researchers in federally funded centers and think tanks creates many problems, including a loss of useful information and bad public policies based on skewed information.

But, two adverse consequences worry me the most. First, these badly behaved researchers are the only ones that most journalists and policy-makers pay any attention to.

Second, the effects of their bad behavior are 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 of which comes from CRESST, and imply that the vast majority of relevant research does not exist.⁷

More recently, the Organisation for Economic Co-operation and Development (OECD) published a completely one-sided study on educational assessment that ignores most of the relevant research literature and highlights that conducted at a certain US federal research center and several US think tanks (Phelps 2013, 2014). Their skewed recommendations are now the world's.

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, MA: 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, DC: National Research Council.
- Chingos, M. (2012, November). *Strength in numbers: State spending on K-12 assessment systems*. Washington, DC: Brookings Institution. Retrieved March 12, 2014 from: <http://www.brookings.edu/~media/research/files/reports/2012/11>

⁷ See Clarke 2013, Koretz 2013, & 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, first with its affiliated researchers and graduates Thomas Kelleghan, then Vincent Greaney, and now Marguerite Clarke, all Irish citizens. Leadership succession in this office of the World Bank is not meritocratic; it is filial.

[/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, DC.

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 & H.J. Walberg (Eds.), *School Accountability*, Stanford, CA: 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, DC.

Monk, D.H. (1995, Spring). The costs of pupil performance assessment: A summary report, *Journal of Education Finance*, 20(4), pp. 363–371.

Monk, D.H. (1996, Spring). The importance of balance in the study of educational costs, *Journal of Education Finance*, 21(4), pp. 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, DC: American Federation of Teachers.

Phelps, R.P. (1996, Spring). Mis-conceptualizing the costs of large-scale assessment, *Journal of Education Finance*, 21(4), pp. 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, AL.

- 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, DC: 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, forthcoming). A review of *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*. *Assessment in Education: Principles, Policy, and Practice*.
- Picus, L.O., & 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, DC.

U.S. GAO. (1993, January). *Student testing: Current extent and expenditures, with cost estimates for a national examination*. GAO/PEMD-93-8. Washington, DC: US General Accounting Office.