

Dismissive Reviews: Academe's Memory Hole

Richard P. Phelps

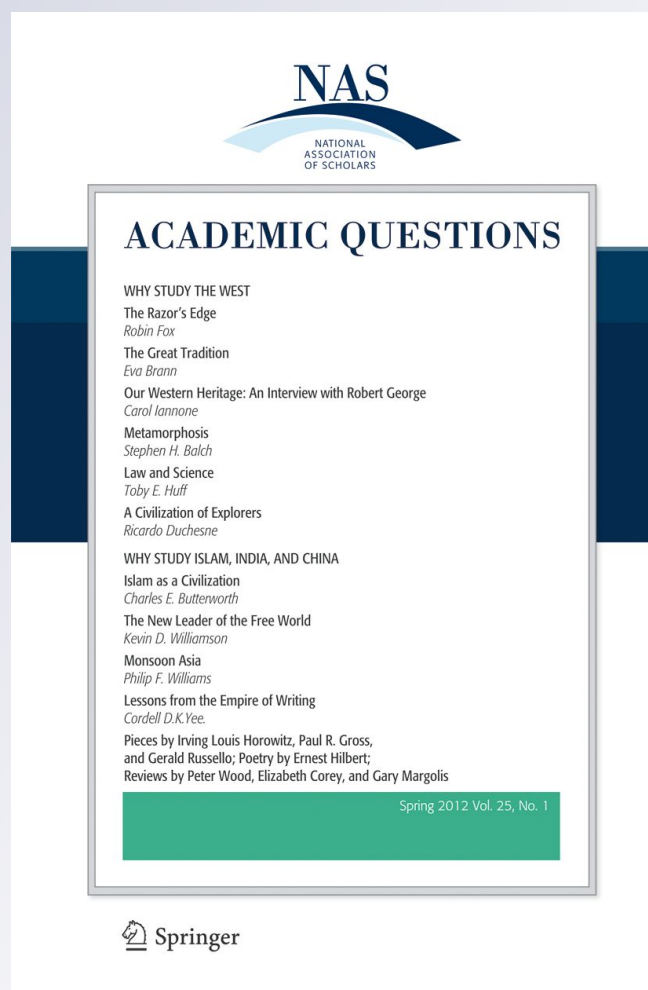
Academic Questions

A Publication of the National
Association of Scholars

ISSN 0895-4852

Acad. Quest.

DOI 10.1007/s12129-012-9289-4



Your article is protected by copyright and all rights are held exclusively by The Author(s). This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.

Dismissive Reviews: Academe's Memory Hole

Richard P. Phelps

© The Author(s) 2012. This article is published with open access at Springerlink.com

In scholarly terms, a *review of the literature* or *literature review* is a summation of the previous research that has been done on a particular topic. With a *dismissive literature review*, a researcher assures the public that no one has yet studied a topic or that very little has been done on it. A *firstness claim* is a particular type of dismissive review in which a researcher insists that he is the first to study a topic. Of course, firstness claims and dismissive reviews can be accurate—for example, with genuinely new scientific discoveries or technical inventions. But that does not explain their prevalence in nonscientific, nontechnical fields, such as education, economics, and public policy, nor does it explain their sheer abundance across all fields.

See for yourself. Access a database that allows searching by phrases (e.g., Google, Yahoo Search, Bing) and try some of these: “this is the first study,” “no previous studies,” “paucity of research,” “there have been no studies,” “few studies,” “little research,” or their variations. When I first tried this, I expected hundreds of hits; I got hundreds of thousands.

Granted, the search “counts” in some of the most popular search engines are rough estimates and not actual counts. Still, scrolling through, say, just the first five hundred results from one of these searches can be

Richard P. Phelps is founder of the *Nonpartisan Education Review* (<http://www.nonpartisaneducation.org>); mail@nonpartisaneducation.org. His most recent books are *Defending Standardized Testing* (Psychology Press, 2005), *Standardized Testing Primer* (Peter Lang, 2007), and *Correcting Fallacies about Educational and Psychological Testing* (American Psychological Association, 2009).

revealing—firstness claims and dismissive reviews are far more common than they have any right to be.

And, when false, they can be harmful. Dismissive reviews assure readers that no other research has been conducted on a topic, *ergo*, there is no reason to look for it. Perhaps it would be okay if everyone knew that most dismissive reviews were bunk and so discounted all of them. But, unfortunately, many believe them, and reinforce the harm by spreading them. The laziest dismissive review is one that merely references someone else's.

Dismissive reviews aren't just lazy, though; they are gratuitous. If there ever was a requirement that each and every research article must include a thorough literature review, it has long since lapsed in most journals. Scholars do not need to hide that they have not searched everywhere they could. Research has accumulated in many fields to such a volume that familiarity with an entire literature would now be too time-consuming for any individual. In most fields, when someone writes a dismissive review and claims command of an entire research literature, they claim a near impossible accomplishment.

Sure, digitalization and the Internet have made it easier to search for research work. But a countertrend of proliferation—of research and research categories, methods, vocabulary, institutions, and dissemination outlets—has coincidentally made it more difficult.

In 2008, 2.5 million Ph.D.'s resided in the United States alone. They, and others like them now fill more than 9,100 journals for over 2,200 publishers in approximately 230 disciplines from 78 countries, according to *Journal Citation Reports*.¹ *ProQuest Dissertations and Theses—Full Text* “includes 2.7 million searchable citations to dissertations and theses from around the world from 1861 to the present day....More than 70,000 new full text dissertations and theses are added to the database each year.”² Ulrich's alone covers the publication in 2012 of more than 300,000 serials, from over 90,000 publishers, in more than 950 subject areas, and over 200 languages.³

¹Thompson Reuters, *Journal Citation Reports*, “JCR Fact Sheet 2010,” <http://wokinfo.com/media/pdf/jcrwebfs.pdf>.

²ProQuest, “ProQuest Dissertations & Theses Database,” <http://www.proquest.com/en-US/catalogs/databases/detail/pqdt.shtml>.

³SerialsSolutions, “Ulrich's,” <http://www.serialssolutions.com/en/services/ulrichs>.

Ironically, as research studies accumulate, so do the incentives and opportunities to dismiss large numbers of them. Perhaps we can empathize with an impoverished Ph.D. student cutting corners to meet a dissertation deadline. In the research fields I know best, however, dismissive reviews are popular with some of the most celebrated and rewarded scholars working at the most elite institutions. Indeed, some of these well-known scholars are “serial dismissers,” repeatedly asserting the nonexistence of previous research in more than a few of their articles across more than a few different topics.

As a cynic might ask, why shouldn't they? Professional rewards accrue to “pioneering work” and, to my observation, there are no punishments for dismissive reviews. Even if exposed, a dismissive reviewer can always fall back on the “I didn't know” excuse.

By contrast, accusing another scholar of falsely dismissing an extant research literature poses considerable risk. The accuser might be labeled unprofessional for criticizing a highly-regarded scholar for a presumably honest mistake. Indeed, I have been so accused.⁴ Yet, most of those I have criticized for dismissive reviewing *had been* directly informed of an extant research literature—I told them—and still dismissed it, suggesting willfulness. Other dismissive reviewers have asserted the nonexistence of a research literature a century old and several hundred studies thick. When someone claims to have looked but was unable to find trees in a forest that large, can we not assume that individual is lying—at least about having looked?

Whereas rich professional rewards await those considered to be the first to study a topic, conducting a top-notch, high-quality literature review bestows none. After all, it isn't “original work.” (Note also which of the two activities is more likely to be called a “contribution” to scholarship.) In addition, there are substantial opportunity costs. Thorough reviews demand a huge investment of time—one that grows larger with the accumulation of each new journal issue. In a publish-or-perish environment, really reviewing the research literature before presenting one's own research impedes one's professional progress.

⁴Scott O. Lilienfeld and April D. Thames, review of *Correcting Fallacies about Educational and Psychological Testing*, ed. Richard P. Phelps, *Archives of Clinical Neuropsychology* 24, no. 6 (2009): 631–33.

How did it come to this? I tender a few hypotheses:

(1) Manuscript review complacency?

In judging manuscript submissions, many journal reviewers pay no attention to literature review quality (or, the lack thereof), that is, to an author's summation of previous research on the topic. Perhaps they feel that it is not their responsibility. As a result, the standards used to judge a manuscript author's analysis may differ dramatically from those used to judge the literature review component, where convenience samples and hearsay are considered sufficiently rigorous. Ambitious researchers write dismissive reviews early in their careers, learn that reviewers pay no attention, and so keep writing them.

(2) Research Parochialism?

The proliferation of subject fields, subfields, and researcher specializations exacerbates the problem. With time, it becomes more and more difficult for specialists to know even the vocabulary of other fields much less the content. Besides, professional advancement is determined by one's colleagues in the same field. It is professionally beneficial to pay attention to their work on a topic, but not to the work in other disciplines, even when that work may bear on the topic. Furthermore, many—indeed, likely most—scholars do not attempt to read research written in unfamiliar languages.

(3) Winning is everything?

Claiming that others' work does not exist is an easy way to win a debate.

I surmise that dismissive reviews must be more common in some research fields than in others. Research conversations are simply more open in some fields than in others and my field—education research—may be one of the most politicized.

Granted, even in education, all research studies and all viewpoints can be published somewhere. But not all can be published somewhere that matters. The education research literature is massive and inevitably most of it is ignored. The tiny portion influencing policy is that which rises above the “celebrity threshold,” where rich and powerful interests promote their work (think government- and foundation-funded research centers, wealthier universities with dedicated research promotion offices, think tanks, and the like).

The rest is easily dismissed regardless of quality. The vast numbers of researchers operating below the celebrity threshold include not only the many

academics unlucky enough to be left out of one of the highly-promotional groups, but also civil servants—who are restricted from promoting or defending their work—corporate researchers doing proprietary work, and, obviously, the deceased. Live, dead, or undead, producers of work below the celebrity threshold are “zombie researchers.”

This particular zombie researcher, for example, recently completed a meta-analysis and summary of the research literature on the effects of educational testing on student academic achievement published between 1910 and 2010, and anticipates that it will receive little or no attention in celebrity research circles or the media.⁵ Over three thousand documents were reviewed and close to a thousand studies included in the analysis.⁶

It stands to reason that there should be so many studies. Educational standards and standardized tests have existed for decades. Psychologists first developed the “scientific” standardized test over a century ago and they, along with program evaluators and education practitioners, have conducted literally millions of education research studies since.

Nonetheless, over the past couple of decades, a large number of prominent, well-appointed, well-rewarded scholars have repeatedly asserted a dearth of research on the effects of standardized testing and, in particular, testing with stakes (i.e., consequences)—sometimes called “test-based accountability.”

“[I]t is important to keep in mind the limited body of data on the subject. We are just getting started in terms of solid research on standards, testing and accountability,” said Tom Loveless, Harvard professor and Brookings Institution education policy expert in 2003.⁷ “Debates over accountability

⁵Richard P. Phelps, “The Effect of Testing on Achievement: Meta-Analyses and Research Summary, 1910–2010: Source List, Effect Sizes, and References for Quantitative Studies,” *Nonpartisan Education Review* 7, no. 2 (2011): 1–25, <http://www.nonpartisaneducation.org/Review/Resources/QuantitativeList.htm>; “The Effect of Testing on Achievement: Meta-Analyses and Research Summary, 1910–2010: Source List, Effect Sizes, and References for Survey Studies,” *Nonpartisan Education Review* 7, no. 3 (2011): 1–23, <http://www.nonpartisaneducation.org/Review/Resources/SurveyList.htm>; “The Effect of Testing on Achievement: Meta-Analyses and Research Summary, 1910–2010: Source List, Effect Sizes, and References for Qualitative Studies,” *Nonpartisan Education Review* 7, no. 4 (2011): 1–30, <http://www.nonpartisaneducation.org/Review/Resources/QualitativeList.htm>; and “The Effect of Testing on Student Achievement, 1910–2010,” *International Journal of Testing* 12, no. 1 (2012): 21–43.

⁶An accounting of the entire literature, however, is far from complete. I have been searching for and collecting studies on this topic for over a decade, with substantial help from librarians. Yet, I still often encounter studies or mentions of possibly eligible studies that I’ve missed, despite having already spent thousands of hours looking and reviewing. (Indeed, should anyone be interested in funding my time, I would be happy to review the several hundred additional studies I have found to date, and recalculate my meta-analytic results.)

⁷Tom Loveless, quoted in “New Report Confirms,” U.S. Congress: Committee on Education and the Workforce, news release, February 11, 2003.

are sorely lacking in empirical measures of what is actually transpiring,” added Frederick M. Hess, scholar at the American Enterprise Institute and author of “many books.”⁸

“Most of the evidence is unpublished at this point, and the answers that exist are ‘partial’ at best,” offered Erik Hanushek, a Republican Party advisor and Stanford University and Hoover Institution economist in 2002.⁹ In a remarkable moment of irony, Hanushek, who casually dismisses the work of so many others, is quoted as saying, “Some academics are so eager to step out on policy issues that they don’t bother to find out what the reality is.”¹⁰

Daniel Koretz, a Harvard University professor and longtime associate of the federally-funded Center for Research on Evaluation, Educational Standards, and Student Testing (CRESST), wrote in 1996, “Despite the long history of assessment-based accountability, hard evidence about its effects is surprisingly sparse, and the little evidence that is available is not encouraging.”¹¹ That same

⁸Frederick M. Hess, “Commentary: Accountability Policy and Scholarly Research,” *Educational Measurement: Issues and Practice* 24, no. 4 (December 2005): 57. For example, the mastery learning/mastery testing experiments conducted from the 1960s through today varied incentives, frequency of tests, types of tests, and many other factors to determine the optimal structure of testing programs. Researchers included such notables as Bloom, Carroll, Keller, Block, Burns, Wentling, Anderson, Hymel, Kulik, Tierney, Cross, Okey, Guskey, Gates, and Jones.

The vast literature on effective schools dates back a half-century and arrives at remarkably uniform conclusions about what works to make schools effective—goal-setting, high standards, and frequent testing. Researchers have included Levine, Lezotte, Cotton, Purkey, Smith, Kiemig, Good, Grouws, Wildemuth, Rutter, Taylor, Valentine, Jones, Clark, Lotto, and Astuto.

International organizations such as the World Bank or the Asian Development Bank have studied the effects of testing on education programs they sponsor. Researchers have included Somerset, Heynemann, Ransom, Psacharopoulos, Velez, Brooke, Oxenham, Bude, Chapman, Snyder, and Pronaratna.

See Richard M. Phelps, “The Rich, Robust Research Literature on Testing’s Achievement Benefits,” in Richard P. Phelps, ed., *Defending Standardized Testing* (Mahwah, NJ: Lawrence Erlbaum, 2005), 55–90, app. B; “Educational Achievement Testing: Critiques and Rebuttals,” in Richard P. Phelps, ed., *Correcting Fallacies about Educational and Psychological Testing* (Washington, DC: American Psychological Association, 2008), 89–146, app. C, D; “Effect of Testing: Quantitative Studies”; “Effect of Testing: Survey Studies”; “Effect of Testing: Qualitative Studies”; and “Effect of Testing on Student Achievement, 1910–2010.”

⁹Eric A. Hanushek and Margaret E. Raymond, “Lessons about the Design of State Accountability Systems,” paper prepared for Taking Account of Accountability: Assessing Policy and Politics, Harvard University, Cambridge, MA, June 9–11, 2002; “Improving Educational Quality: How Best to Evaluate Our Schools?” paper prepared for Education in the 21st Century: Meeting the Challenges of a Changing World, Federal Reserve Bank of Boston, June 2002; “Lessons about the Design of State Accountability Systems,” in *No Child Left Behind? The Politics and Practice of Accountability*, ed. Paul E. Peterson and Martin R. West (Washington, DC: Brookings Institution, 2003), 126–51. Lynn Olson, “Accountability Studies Find Mixed Impact on Achievement,” *Education Week*, June 19, 2002, 13.

¹⁰Rick Hess, “Professor Pallas’s Inept, Irresponsible Attack on DCPS,” *Education Week on the Web*, August 2, 2010, http://blogs.edweek.org/edweek/rick_hess_straight_up/2010/08/professor_pallass_inept_irresponsible_attack_on_dcps.html.

¹¹Daniel Koretz, “Using Student Assessments for Educational Accountability,” in *Improving America’s Schools: The Role of Incentives*, ed. Eric A. Hanushek and Dale W. Jorgenson (Washington, DC: National Academy Press, 1996).

year Stanford professor Sean Reardon, reporting his own work on the topic, said, "Virtually no evidence exists about the merits or flaws of MCTs [minimum competency tests]." Reardon has since received over \$10 million in research funding from the federal government and foundations.¹²

In 2002, Brian A. Jacob, who has taught at Harvard and the University of Michigan, asserted: "Despite its increasing popularity within education, there is little empirical evidence on test-based accountability (also referred to as high-stakes testing)."¹³ A year earlier, Jacob wrote: "[T]he debate surrounding [minimum competency tests] remains much the same, consisting primarily of opinion and speculation.... A lack of solid empirical research."¹⁴ About the same time, Jacob and Steven Levitt, co-author of the best-selling *Freakonomics* (Levitt & Dubner, 2005), claimed to have conducted the "first systematic attempt to (1) identify the overall prevalence of teacher cheating [in standardized test administrations] and (2) analyze the factors that predict cheating."¹⁵

In 2008, Jacob won the David N. Kershaw Award, "established to honor persons who, at under the age of 40, have made a distinguished contribution to the field of public policy analysis and management....[T]he award consists of a commemorative medal and cash prize of \$10,000 [and is] among the largest awards made to recognize contributions related to public policy and social science."¹⁶

¹²Sean F. Reardon, "Eighth Grade Minimum Competency Testing and Early High School Dropout Patterns" (paper, American Educational Research Association annual meeting, New York, NY, April 8, 1996). The many studies of district and state minimum competency or diploma testing programs popular from the 1960s through the 1980s were conducted by Fincher, Jackson, Battiste, Corcoran, Jacobsen, Tanner, Boylan, Saxon, Anderson, Muir, Bateson, Blackmore, Rogers, Zigarelli, Schafer, Hultgren, Hawley, Abrams, Seubert, Mazzone, Brookhart, Mendro, Herrick, Webster, Orsack, Weerasinghe, and Bembry. See Phelps, "Rich, Robust Research Literature," "Educational Achievement Testing," "Effect of Testing: Quantitative Studies," "Effect of Testing: Survey Studies," "Effect of Testing: Qualitative Studies," and "Effect of Testing on Student Achievement, 1910–2010."

¹³Brian A. Jacob, "Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools" (NBER Working Paper No. W8968, National Bureau of Economic Research, Cambridge, MA, 2002), 2.

¹⁴Brian A. Jacob, "Getting Tough? The Impact of High School Graduation Exams," *Educational Evaluation and Policy Analysis* 23, no. 2 (Fall 2001): 334.

¹⁵Brian A. Jacob and Steven D. Levitt, "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating," *Quarterly Journal of Economics* (August 2003): 845, <http://pricetheory.uchicago.edu/levitt/Papers/JacobLevitt2003.pdf>. Since before 1960 test publishers have analyzed classroom-level variations in each test administration, looking for evidence of teacher or student cheating. This has transpired tens of thousands, perhaps hundreds of thousands, of times.

¹⁶University of Michigan, Gerald R. Ford School of Public Policy, "Brian Jacob Earns Prestigious David N. Kershaw Award and Prize," faculty news, October 7, 2008, http://www.fordschool.umich.edu/news/?news_id=87.

In 2002, Jacob co-wrote a study with Anthony Bryk, the president of the Carnegie Foundation for the Advancement of Teaching, in which they claimed to have studied “one of the first large, urban school districts to implement high-stakes testing” in the late 1990s.¹⁷ (In fact, U.S. school districts have hosted comprehensive high-stakes testing programs by the hundreds, and for over a hundred years.) Brian Jacob alone has declared the nonexistence of the good work of perhaps over a thousand scholars, living and deceased, in the United States and the rest of the world, and has been rewarded for it.

Dismissive reviews abound among related education research topics, too. Consider this 1993 claim from Robert Linn, an individual some consider to be the nation’s foremost testing expert.: “[T]here has been surprisingly little empirical research on the effects of different motivation conditions on test performance. Before examining the paucity of research on the relationship of motivation and test performance...”¹⁸ Gregory Cizek, a Republican Party education policy advisor and current president of the National Council on Measurement in Education—the primary association for those working in educational testing—agreed in 2001: “[T]he evidence regarding the effects of large-scale assessments on teacher motivation...is sketchy; and with respect to assessment impacts on the affect of students, we are again in a subarea where there is not a great deal of empirical evidence.”¹⁹ Are they correct? Not even close.²⁰

Or consider this from Todd R. Stinebrickner and Ralph Stinebrickner of the University of Western Ontario in 2007: “Despite the large amount of attention that has been paid recently to understanding the determinants of educational outcomes, knowledge of the causal effect of the most

¹⁷Melissa Roderick, Brian A. Jacob, and Anthony S. Bryk, “The Impact of High-Stakes Testing in Chicago on Student Achievement in the Promotional Gate Grades,” *Educational Evaluation and Policy Analysis* 24, no. 4 (Winter 2002): 333–57. Brian A. Jacob, “High Stakes in Chicago,” *Education Next* 3, no. 1 (Winter 2003): 66, <http://educationnext.org/highstakesinchicago/>.

¹⁸Vonda L. Kiplinger and Robert L. Linn, *Raising the Stakes of Test Administration: The Impact on Student Performance on NAEP*, CSE Technical Report 360 (Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, 1993), <http://www.cse.ucla.edu/products/reports/TECH360.pdf>.

¹⁹Gregory J. Cizek, “More Unintended Consequences of High-Stakes Testing?” *Educational Measurement: Issues and Practice* 20, no. 4 (Winter 2001): 19–28.

²⁰Many researchers studied the role of motivation in educational testing before the year 2000. They have included Tyler, Anderson, Kulik & Kulik, Crooks, O’Leary, Drabman, Kazdin, Bootzin, Staats, Resnick, Covington, Brown, Walberg, Pressey, Wood, Olmsted, Chen, Stevenson, and Resnick. See Phelps, “Rich, Robust Research Literature,” “Educational Achievement Testing,” “Effect of Testing: Quantitative Studies,” “Effect of Testing: Survey Studies,” “Effect of Testing: Qualitative Studies,” and “Effect of Testing on Student Achievement, 1910–2010.”

fundamental input...—student study time and effort—has remained virtually non-existent. In this paper...”²¹

Being the first, apparently, to take on such an obvious topic for study won the Stinebrickners the Kenneth J. Arrow Prize in Economic Analysis & Policy, “for making an outstanding contribution to economics.” The award carries a \$5,000 honorarium and publication in a journal “that accepts less than 1% of all submissions.”²²

In 2005, Stanford professor Eric Bettinger and Harvard professor Bridgett Terry Long, who have received tens of millions of dollars in research grants and dozens of honors from the most distinguished national research groups and foundations, asserted that “Despite the growing debate and the thousands of under prepared students who enter college each year, there is almost no research on the impact of remediation on student outcomes. This project addresses this critical issue...”²³ Almost no research? Hardly.²⁴

In 2000, David Figlio, a Northwestern University professor, recipient of countless awards and research grants, and “referee of approximately 60 papers and proposals per year for over 30 journals, Federal agencies, and private foundations,” wrote: “While high standards have been advocated by policy-makers...very little is known about their effects on outcomes....This paper provides the first empirical evidence on the effects of grading standards, measured at the teacher level.”²⁵ The same year, one of the nation's foremost scholars in program evaluation—the late Frederick

²¹Stinebrickner and Stinebrickner measured “studying” by comparing the academic outcomes of college dormitory students with access to video game players to those without. Todd R. Stinebrickner and Ralph Stinebrickner, “The Causal Effect of Studying on Academic Performance” (NBER Working Paper No. 13341, National Bureau of Economic Research, Cambridge, MA, 2007), <http://www.nber.org/papers/w13341>; “The Causal Effect of Studying on Academic Performance,” *B.E. Journal of Economic Analysis & Policy* 8, no. 1 (June 2008): n.p.

²²Berkeley Electronic Press, Kenneth J. Arrow Prizes in Theoretical Economics, Macroeconomics, and Economic Analysis & Policy, <http://www.bepress.com/press/13/>.

²³Eric P. Bettinger and Bridgett Terry Long Bettinger, “Addressing the Needs of Under-Prepared Students in Higher Education: Does College Remediation Work?” (NBER Working Paper No. 13325, National Bureau of Economic Research, Cambridge, MA, 2005), <http://www.nber.org/papers/w13325>.

²⁴Developmental (i.e., remedial) education researchers had conducted many studies prior to 2000 to determine what works best to keep students from failing in their “courses of last resort,” after which there are no alternatives. Researchers have included Boylan, Roueche, McCabe, Wheeler, Kulik, Bonham, Claxton, Bliss, Schonecker, Chen, Chang, and Kirk. See Phelps, “Rich, Robust Research Literature,” “Educational Achievement Testing,” “Effect of Testing: Quantitative Studies,” “Effect of Testing: Survey Studies,” “Effect of Testing: Qualitative Studies,” and “Effect of Testing on Student Achievement, 1910–2010.”

²⁵David N. Figlio and Maurice E. Lucas, “Do High Grading Standards Affect Student Performance?” (NBER Working Paper No. W7985, National Bureau of Economic Research, Cambridge, MA, 2000), <http://www.nber.org/papers/w7985>.

Mosteller, who received dozens of honors, including the titling of the Campbell Collaborative's Frederick Mosteller Award for Distinctive Contributions to Systematic Reviewing—co-wrote, "Little empirical evidence supports or refutes the existence of a causal link between standards and enhanced student learning. ...[W]e found few empirical studies of the impact...of standards on schools and students."²⁶ Another scholar with an award named after him, Robert Linn, wrote, "Too little attention has been given to the evaluation of the alignment of assessments and standards."²⁷ Believe them? You shouldn't.²⁸

In 1999, Helen Ladd, a Duke University professor, Harvard Ph.D., Democratic Party advisor, author of fourteen books, forty-two book chapters, and hundreds of journal articles and reports, wrote: "Given the widespread interest in school-based recognition and reward programs, it is surprising how little evaluation has been done of their impacts.... This paper provides one of the few evaluations of the effects of such programs on student outcomes."²⁹ Little evaluation has been done? Not really.³⁰

In 2006, think-tanker Frederick Hess claimed, "Despite the importance of arbitration [in education labor negotiations], the process has largely escaped either scholarly or journalistic attention" even as he himself wrote on the topic.³¹ Believe it? Me neither.³²

²⁶The Campbell Collaboration, "The Frederick Mosteller Award," http://www.campbellcollaboration.org/c2_awards/frederick_mosteller_award.php. Bill Nave, Edward Miech, and Frederick Mosteller, "A Lapse in Standards: Linking Standards-Based Reform with Student Achievement," *Phi Delta Kappan* 82, no. 2 (October 2000): 128–32.

²⁷Robert L. Linn, *Issues in the Design of Accountability Systems*, CRESST Report 650 (Los Angeles, CA: Center for Research on Educational Standards and Student Testing, 2005).

²⁸Studies dating back to the 1910s cover the effect of goal setting, standards and alignment on teachers, instruction, and student learning. The researchers involved have included Tyler, Panlasigui, Knight, Resnick, Robinson, Thomas, Stark, Shaw, Lowther, Csikszentmihalyi, Pine, Pomplun, Fontana, Natriello, Dornbusch, Kasdin, Bootzin, Chaney, and Burgdorf. See Phelps, "Rich, Robust Research Literature," "Educational Achievement Testing," "Effect of Testing: Quantitative Studies," "Effect of Testing: Survey Studies," "Effect of Testing: Qualitative Studies," and "Effect of Testing on Student Achievement, 1910–2010."

²⁹Helen F. Ladd, "The Dallas School Accountability and Incentive Program: An Evaluation of Its Impacts on Student Outcomes," *Economics of Education Review* 18, no. 1 (February 1999): 1–16.

³⁰Some of the researchers who, prior to 2000, studied test-based incentive programs include Homme, Csanyi, Gonzales, Rechs, O'Leary, Drabman, Kaszdin, Bootzin, Staats, Cameron, Pierce, McMillan, Corcoran, Roueche, Kirk, Wheeler, Boylan, and Wilson. See Phelps, "Rich, Robust Research Literature," "Educational Achievement Testing," "Effect of Testing: Quantitative Studies," "Effect of Testing: Survey Studies," "Effect of Testing: Qualitative Studies," "Effect of Testing on Student Achievement, 1910–2010."

³¹Frederick M. Hess and Andrew P. Kelley, "Scapegoat, Albatross, or What?" in *Collective Bargaining in Education: Negotiating Change in Today's Schools*, ed. Jane Hannaway and Andrew J. Rotherham (Cambridge, MA: Harvard Educational Publishing Group, 2006), 85.

³²Amazed by the audacity of this claim, Myron Lieberman identified a 1,336-item bibliography published on the topic in 1994 as well as many publications on the topic produced by national associations. *The Educational Morass* (Lanham, MD: Rowan & Littlefield Education, 2007), 287–92.

A review of the lengthy curriculum vitae of some of these dismissive reviewers, with their superabundance of honors, awards, grants, and publications suggests two conclusions:

- They are much too busy to spend time on thorough literature reviews
- Most of them claim a numbingly large volume of scholarly production

Indeed, these reviewers claim so much scholarship it begs the question why they might feel motivated to seek more attention by dismissing others' work. They may be the scholarly equivalent of billionaires for whom no amount of wealth is enough to satisfy. Then again, perhaps they have achieved celebrity status in part because they have been willing to scratch for every little bit of credit throughout their careers.

But boastfulness is not the only problem. None of the dismissive reviews mentioned above pertain to purely academic debates—all pertain to important public policies. Each boast dismissed a research literature relevant to a public need. In some cases, a highly-influential scholar promoted his single work on a topic to the exclusion of *hundreds* of other works conducted by lesser-knowns and the dear departed.

All the aforementioned statements dismissing the research on educational testing were uttered within several years of the 2000 presidential campaign, the only national election in our country's history in which standardized testing was a major campaign issue. Thus, while the most far-reaching federal intervention in U.S. assessment policy—contained inside the No Child Left Behind (NCLB) Act—was being considered, the most influential research advisors for both major political parties managed to convince policy makers that no research existed to help guide them in their program design. That casually, a century's worth of relevant research was declared nonexistent. The result? The research-uninformed NCLB Act. New research was needed to fill the void, according to some of our nation's premier scholars, and they were willing to do it, for a fee.

Prior research and experience would have told policy makers that most of the motivational benefits of standardized tests required consequences for the students and not just for the schools. Those stakes needn't be very high to be effective, but there must be some. As NCLB imposes stakes on schools, but not on students, who knows if the students even try to perform well?

Prior research and experience would have informed policy makers that educators are intelligent people who respond to incentives, and who will game a system if they are given an opportunity to do so. The NCLB Act left

many aspects of the test administration process that profoundly affect scores (e.g., incentives and motivation, security, cut scores, curricular alignment) up for grabs and open to manipulation by local and state officials.

Prior research and experience would have informed policy makers that different tests get different results and one should not expect average scores from different tests to rise and fall in unison over time (as some interpreters of the NCLB Act seem to expect with the National Assessment of Educational Progress benchmark).

Prior research and experience would have informed policy makers that the public was not in favor of punishing poorly-performing schools (as NCLB does), but was in favor of applying consequences to poorly-performing students and teachers (which NCLB does not).

The resulting scantily-informed public policy includes a national testing program that would hardly be recognizable anywhere outside of North America. The standardized testing component of NCLB includes no consequences for the students. This sends the subliminal message to the students that they need not work very hard and one of testing's largest potential benefit—motivation—is not even accrued.

By contrast, schools are held accountable for students' test performance; they are held responsible for the behavior of other human beings over whom they have little control. Moreover, the most important potential supporters of testing programs—classroom teachers and school administrators—are alienated, put into the demeaning position of cajoling students to cooperate.

Had the policy makers and planners involved in designing the NCLB Act simply read the freely-available research literature instead of funding expensive new studies and waiting for their few results, they would have received more value for their money, gotten more and better information, and gotten it earlier when they actually needed it.

With the single exception of the federal mandate, there was no aspect of the NCLB accountability initiative that had not been tried and studied before. Every one of the NCLB Act's failings was perfectly predictable, based on decades of prior experience and research. Moreover, there were better alternatives for every characteristic of the program that had also been tried and studied thoroughly by researchers in psychology, education, and program evaluation. Yet, policy makers were made aware of none of them.

The dismissive reviews that misinformed the NCLB policy makers mirrored those made by the National Research Council's Board on Testing

and Assessment in 1999.³³ Perhaps not surprisingly, most of the dismissive reviewers cited above have served on National Research Council (NRC) panels.³⁴

Coincident with my zombie-researcher meta-analysis, another NRC report on standardized testing was published in 2011.³⁵ It generously praises the work of most of the dismissive reviewers cited above, implies that little other work worth considering exists, and reemphasizes the alleged paltry size of the research literature. The timing of the report's release anticipates Congressional consideration of the reauthorization of the NCLB Act.

What Can Be Done?

What can be done about the information suppression resulting from glib dismissive reviews? The situation could be much improved if all scholars were made to review literature in the meta-analyst's way—instead of implying command of an entire research literature, specify exactly where one has looked and summarize only what is found there.

More generally, I believe that we should redefine the meaning of “a contribution” to research. Currently, original works are considered contributions, and quality literature reviews are not. But, what of the scholar who dismisses much of the research literature as nonexistent (or no good) each time he “contributes” an original work? That scholar is subtracting more from society's working memory than adding. That scholar's “value added” is negative.³⁶

³³Jay P. Heubert and Robert M. Hauser, eds., *High Stakes: Testing for Tracking, Promotion, and Graduation* (Washington, DC: National Research Council, 1999).

³⁴I have criticized the NRC's work on testing before and concluded that it has been captured by a particular group of vested interests from the federally-funded CRESST. See Richard P. Phelps, “Education Establishment Bias? A Look at the National Research Council's Critique of Test Utility Studies,” *Industrial-Organizational Psychologist* 36, no. 4 (April 1999): 37–49; review of *High Stakes: Testing for Tracking, Promotion, and Graduation*, ed. Jay P. Heubert and Robert M. Hauser, *Educational and Psychological Measurement* 60, no. 6 (December 2000): 992–99; and “Educational Achievement Testing: Critiques and Rebuttals.” Most of the dismissive reviewers cited above have served on U.S. Department of Education Institute of Education Sciences (IES) advisory boards and technical committees and have received taxpayer funds as members of IES research and development centers. In addition, all of the economists cited above have been affiliated with the National Bureau of Economic Research.

³⁵Michael Hout and Stuart W. Elliot, eds., *Incentives and Test-Based Accountability in Education* (Washington, DC: National Research Council, 2011).

³⁶In a recent review-editorial, *The Economist* magazine ribs doomsayers and hand-wringers, asserting that research is always improving conditions, despite the various impediments of human behavior: “[O]ur authors are certainly right about one thing. Knowledge is cumulative.” If only that were true. “Now for Some Good News: Two Books Argue That the Future Is Brighter Than We Think,” *Schumpeter* (blog), *The Economist*, March 3, 2012, <http://www.economist.com/node/21548937>.

Sadly, it may already be too late to stop the rampant information suppression and our regressive diminution of knowledge. Some researchers seem to have adopted an “everyone does it” rationale. They are now invested in their claims, and some of them lead their disciplines—they are the same people to whom one would normally direct an appeal for ethical reform. It may sound trite, but I believe it to be true: scholars write dismissive reviews because they can. Unless and until dismissive reviews begin to carry some risk, we should expect to continue to see them in abundance.

Even more disturbing, federal funding of research centers apparently provides sufficient money, power, and status to incubate dismissive reviewers, for example at CRESST, headquartered at UCLA. The expenditure of hundreds of millions of taxpayer dollars on these centers is justified by assertions that not enough research exists and more is needed. But, in some cases the net result of taxpayer investment is a diminution of knowledge in exchange for boosting the careers of a few.

With dismissive reviews, society loses information, and that which remains is skewed in favor of those with the resources to promote their own. Public policy decisions are then based on limited and skewed information. And, governments (i.e., taxpayers) and foundations pay again and again for research that has already been done.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.