

# They Blinded Me with Political Science: On the Use of Nonpeer Reviewed Research in Education Policy

Edward Muir, American Federation of Teachers, AFL-CIO<sup>1</sup>

On March 4, 1998, a story appeared in *Education Week* about research by Mark Schneider, Paul Teske, and Christine Roch (hereafter referred to as STR). It summarized their research, which claims that an innovative public school choice program in District 4 in New York City is responsible for raising student achievement (Walsh 1998). This followed a news story in *The New York Times* (Hartcollis 1998). Some weeks later Paul Teske presented the paper at the Midwest Political Science Association. There was a lively discussion of methodology, and the panel ended.

These are relatively mundane political science events, but the order in which they come raises a troubling issue that political scientists need to ponder. STR's results were used to influence the policy debate before they were reviewed by peers either at an academic conference or for publication in a refereed journal.

The primary issue raised by the publication of their latest results in *The New York Times* does not concern the quality of the work, but the appropriate use of unpublished social science research to affect political debate.<sup>2</sup> This is generally not a practice seen in the so-called harder sciences. Instead, the leading journals, such as *Nature* and *Science*, may make efforts to publicize results that will have broader interest, and science reporters scan the leading journals for story ideas. Companies might announce drug discoveries but, before a product can be used

on patients, it has to undergo FDA review. While the announcement of a new cancer treatment might affect stock prices, it won't affect patients until the review procedures, which are intrinsic to our concept of science, are completed. This is not so in political science, where results can be announced to the press and then put into practice by officials without any further review. Since *PS* is the journal for our professional issues, it is the appropriate forum to raise concerns about the practice of providing the political debate with fodder that has not been properly tested.

Of course, STR are not the only political scientists who have revealed their findings to the world via the editorial page or press releases. They are not even the first political scientists to do so on the issue of school choice. On August 14, 1996, Jay Greene and Paul Peterson published "School Choice Data Rescued from Bad Science" on the opinion page of the *Wall Street Journal*. This opinion piece is a summary of a paper, coauthored with Jingtiao Du (the authors are hereafter referred to as GPD), that was then presented at the 1996 Annual Meeting of the American Political Science Association in San Francisco. GPD claimed that this research indicates that school voucher programs raise reading and math scores and could be used to "cut the gap between whites and minorities by one half." Given that a voucher was assumed to result in a savings to the public, this finding made quite an impression on the press and on policymakers.

The subsequent controversies over the GPD voucher research further reveal the perils of using nonpeer-reviewed work. Other unpublished research, by Alex Molnar in this case, shows that GPD's choice of statistical significance test (a 1-tailed

test and a .10 standard) affected their interpretation of the results. Using the conventional two-tailed test and a .05 standard, Molnar (1998) found that the overwhelming majority of GPD's results were not statistically significant. Voucher advocates have retorted that statistical significance is not a cliff; opponents like myself have noted that in this case it appears to be a slippery slope. Rouse (1998a; 1998b) and Witte (1998) have each published peer-reviewed articles contradicting the claims GPD made for a reading advantage for voucher students. Their research also highlighted the problems of attrition in the sample, and pointed out that students from some Milwaukee public schools outperform the voucher students in reading and match them in math. Further, Levin (1998) found that schools participating in the voucher program received as much as \$500 more per comparable student than the Milwaukee Public Schools, even after public school spending on capital was taken into account. This research calls GPD's findings and the resultant policy prescriptions into question. All of this rebutting occurred before a version of the GPD study was actually published more than two years later in a special edition of the journal *Education and Urban Society* (1999).<sup>3</sup>

While questions have been raised about the results, the impact of GPD's 1996 study on public debate has been telling. It has had greater effect than that of any piece published in the *American Political Science Review* at the same time. Between August 13 and August 26, 1996, there were more than thirty reports concerning GPD's findings in newspapers across the country, including *The Washington Times* and *The New York Times*. For example, an editorial in the *Providence Jour-*

---

**Edward Muir** has been an associate in the research department of the American Federation of Teachers since 1997. He has been a member of the American Political Science Association since 1993. He is currently an investigator on the U.S. Department of Education's National Charter School Finance Study.

*nal Bulletin*, which was picked up by at least one other paper, stated: "Peterson and Greene did it right: They compared students in the pilot program with similarly situated youngsters who had applied for the voucher-choice program but had been left out. . . . Officials in other communities ought to keep a close watch on the Milwaukee experiment" ("Maybe Vouchers" 1996). Their work also received radio coverage, particularly on Rush Limbaugh's show, and Lamar Alexander discussed the findings of this new "Harvard Study" on *The Newshour with Jim Lehrer*. Both Greene and Peterson have subsequently testified to Congress on this issue, and Peterson testified in the Wisconsin case concerning the expansion of the voucher program to parochial schools. The state upheld the expansion of the program, and the U.S. Supreme Court has subsequently refused to review the decision. During the election of 1998, the study was cited in campaign literature justifying a tuition tax credit ballot initiative in Colorado.

The popular press proved much less enamored of the articles published in the June 1996 issue of *APSR*, which appeared at about the same time as the release of the GPD report. A Nexus search on the name of each author of the 10 articles revealed that from June 1, 1996 to August 25, 1996, there was no reference to this research in any of the newspapers in the database. While this might be as much of a reflection on the *Review* as on GPD, it still is a telling point. Political scientists have been working diligently towards developing institutional practices regarding the reservation of space in journals for "controversies" or updates on research and requirements concerning the provision of replication data sets with the intent of enhancing the ability of the discipline to police itself and to provide the most rigorous review possible. Appointment to the editorial board of the *APSR* is one of the greatest distinctions a political scientist can attain; these persons are in some sense the gatekeepers of our discipline's integrity. Yet, one has to wonder, given the attention paid to

research published in the *Wall Street Journal* as opposed to the *Review*, if we political scientists are guarding the proper gate.

Early publication of nonpeer-reviewed research is motivated in part by the hunt for grant money. Foundations are often looking to have an immediate impact on public debate. The Century Fund, for example, looks to fund research that yields policy recommendations. STR's research was funded by the Manhattan Institute, which supports school choice. GPD were funded by the Olin Foundation, which is pro-voucher. Funders may place pressure on researchers to publicize their results, or they may be responsible for generating the publicity themselves. Researchers hoping to build a longitudinal data set may release preliminary results in order to generate enthusiasm and funding for further research, thus creating compromises for both researchers and for universities that may find their good name is being rented for use in political debate.

Unfortunately, the practice of using nonpeer-reviewed research is endemic in education policy making. Some education researchers advocate bringing compelling research to policymakers and practitioners as quickly as possible in much the same way that some health-care activists call upon the FDA to make experimental treatments more readily available to patients. In this spirit, the federal government's primary tool for disseminating educational research, the ERIC database, will catalogue just about any research, and contains untold hundreds of nonpeer-reviewed studies.

The result of this culture is an unhappily large number of educational reforms that have been challenged as having little research and data to support them. One analysis of failed reforms that lacked adequate research bases listed the open classroom of the 1970s, multiculturalism, and wholesale inclusion of special education students into regular classrooms (Baines 1997). Another result of this lack of rigor is the promotion of programs that are evaluated solely by their creators. A

Government Accounting Office study of five "promising" school safety initiatives, for instance, included one that was evaluated by the contractor and another that was evaluated by the city agency that designed the program (GAO 1995).

In fact, an education labor-management coalition (including my employer, the American Federation of Teachers) recently felt it necessary to contract with the American Institutes of Research to construct a framework for assessing the models of schoolwide reform currently being sold for use in America's public and private schools. This was done in an effort to fill the gap caused by the lack of proper standards for review of educational research by the academic and policy communities. The framework rates the evaluative support for 24 widely used programs on the basis of criteria such as scope, objectivity of measurement instruments, construct validity, internal validity, sample bias, use of appropriate statistical technique, and external validity, in order to create a methodological validity rating. Out of the two dozen programs examined, only three had both an adequate research base and strong evidence of success. These are in use in 2,200 schools nationwide. Over 6,000 other schools are using the 21 programs with either less rigorous research bases or with limited evidence of success (Herman 1999).<sup>4</sup> Of the 116 studies evaluated by the AIR, many of which have been used by advocates to sell their programs, 66 were done by developers and only 33 of the 116 studies had been published in a peer-reviewed academic journal.

This is an example of low standards for review that political scientists should not be seeking to emulate. However, authors cannot and should not be restrained from participating in the public sphere. Indeed, authors should be encouraged to make their research understandable to a broader public. The duties implicit in being a social scientist do not abrogate the rights or duties of citizenship. Furthermore, there is little that the APSA can do to enforce scientific norms. Having said

that, we as social scientists must find a way to insure that the integrity of our discipline is maintained. Perhaps our best tool is moral suasion.

It is now several decades since Karl Popper first warned, in *The Poverty of Historicism*, against using the trappings of science as totems of competence in political debate. Popper saw this as the creation of pseudoscience, which could be put

to whatever political end its masters desired, at the expense of the truth. In fact, the application of the sobriquet "scientific" creates, in the general audience, a vision of research that has been tested and that is beyond reproach. As a community of scholars, we have developed a number of norms and conventions that are an integral part of distinguishing good social science from pseudo-

science. To present to the general public research that has not endured the scrutiny of peer review and whose statistical results have not been held to the standards established by the social scientific community, while all the time calling the work "political science," is a challenge to the very nature of our enterprise as a community of scholars and citizens.

---

## Notes

1. The views expressed in this article are those of the author and not of the American Federation of Teachers.

2. In fact, STR and other coauthors have already used versions of this data to publish studies in major journals. These include Schneider, Teske, Marschall, Mintrom, and Roch (1997), Schneider, Marschall, Teske,

and Roch (1997), and Schneider, Teske, Marschall, and Roch (1998).

3. This version included some additional adjustments to the data and method of analysis, finding improvements not only in the third and fourth year of the study, as in the first version, but in the first and second year as well. It also claimed that these

achievement effects were produced at lower costs.

4. This is not to say that these programs are bad. In fact, many were rated as "promising" by the evaluators. It is to be hoped that further rigorous research will be done in this area.

---

## References

- Baines, Lawrence. 1997. "Future Schlock . . . Using Fabricated Data and Politically Correct Platitudes in the Name of Educational Reform." *Phi Delta Kappan Magazine* 78(7): 492-98.
- General Accounting Office (GAO). 1995. *School Safety: Promising Initiatives for Addressing School Violence* (HEHS-95-106). Washington, DC: GPO.
- Greene, Jay, Paul Peterson, and Jingtao Du. 1996. "The Effectiveness of School Choice in Milwaukee: A Secondary Analysis of Data from the Program's Evaluation." Presented at the Annual Meeting of the American Political Science Association, San Francisco.
- . 1999. "Effectiveness of School Choice: The Milwaukee Experiment." *Education and Urban Society* 31(January): 190-213.
- Harteollis, Anemona. 1998. "Choice System Helps Schools in East Harlem: Study Says Scores Rose Throughout the District." *The New York Times*, February 24, B5.
- Herman, Rebecca. 1999. *An Educators' Guide to Schoolwide Reform*. Arlington, VA: Educational Research Service.
- Levin, Henry. 1998. "Educational Vouchers: Effectiveness, Choice and Costs." *Journal of Policy Analysis and Management* 17(3): 373-92.
- "Maybe Vouchers Work." 1996. *Providence Journal Bulletin*, August 20, 7.
- Molnar, Alex. 1998. "Smaller Classes, Not Vouchers, Increase Student Achievement." Harrisburg, PA: Keystone Research Center. Manuscript.
- Peterson, Paul, and Jay Greene. 1996. "School Choice Data Rescued from Bad Science." *The Wall Street Journal*, August 14.
- Popper, Karl. 1961. *The Poverty of Historicism*. London: Routledge.
- Rouse, Celia. 1998a. "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program." *The Quarterly Journal of Economics* 113(May): 553-602.
- . 1998b. "Schools and Student Achievement: More Evidence from the Milwaukee Parental Choice Program." *Economic Policy Review* 4(March): 61-76.
- Schneider, Mark, Paul Teske, Melissa Marschall, Michael Mintrom, and Christine Roch. 1997. "Institutional Arrangements and the Creation of Social Capital: The Effects of Public School Choice." *American Political Science Review* 91(March): 82-93.
- Schneider, Mark, Melissa Marschall, Paul Teske, and Christine Roch. 1997. "Networks to Nowhere: Segregation and Stratification in Networks of Information about Schools." *American Journal of Political Science* 41:1201-23.
- Schneider, Mark, Paul Teske, Melissa Marschall, and Christine Roch. 1998. "Shopping for Schools: In the Land of the Blind, The One-Eyed Parent May Be Enough." *American Journal of Political Science* 42(July): 769-93.
- Walsh, Mark. 1998. "Study Credits Choice with Raising Test Scores." *Education Week*, March 4, 3.
- Witte, John. 1998. "The Milwaukee Voucher Experiment." *Educational Evaluation and Policy Analysis* 20(4): 229-51.