

Education Week

Commentary

Doing No Harm

Is It Time for a Research Boycott of High-Stakes Testing?

By James H. Nehring



It is now widely accepted across the scholarly community that the persistent use of high-stakes testing in public education is a potentially harmful practice. The large and growing body of evidence is synthesized nicely in recent publications by Audrey Amrein and David Berliner; Deborah Meier and colleagues; and Wayne Au. In addition, the opinions of leading scholars from a range of research traditions and political perspectives are appearing in the press with increasing frequency. Most notably, Chester E. Finn Jr. and Diane Ravitch, long associated with the politically conservative impulse that spawned the current testing movement, declared in an opinion piece in *The Wall Street Journal* last summer: “We’re already at risk of turning U.S. schools into test-prepping skill factories where nothing matters except exam scores on basic subjects. That’s not what America needs nor is it a sufficient conception of educational accountability.”

Joining these voices, the leading professional organizations in the field have for several years maintained formal positions on the use and abuse of high-stakes testing. From the American Educational Research Association’s statement adopted in 2000: “Decisions that affect individual students’ life chances or educational opportunities should not be made on the basis of test scores alone.” From the American Psychological Association’s statement adopted in 2001: “[W]hen test results are used inappropriately or as a single measure of performance, they can have unintended adverse consequences.”

It is rare for scholars to reach such broad consensus. That it has been achieved on a matter of such immediacy and consequence raises an important question: If the persistent use of high-stakes testing in public schools is a potentially harmful practice, do researchers who draw on the results of such testing legitimize the practice and become complicit in its perpetuation?

To put this question in a larger context, consider the legacy of international agreements that have drawn guidelines for research involving human subjects. Since the Nuremberg Code was established in 1947 following the war-crimes trials in that German city, research involving human subjects has been addressed in the Declaration of Helsinki (1964), the Belmont Report (1979), and the Federal Policy for the Protection of Human Subjects, or “Common Rule” (1991), among other documents. Of particular ethical concern to the drafters of these reports was the welfare of vulnerable populations, including persons under the age of 18.

If, as many have contended, high-stakes tests are potentially harmful to large numbers of children, doesn't it follow, from an ethical standpoint, that the use of the results of these tests in scholarly research is, at best, somewhat questionable?

A test is high-stakes if it is being used to determine, as the AERA puts it, a person's "life chances or educational opportunities." This would include most state-sponsored exit exams. Institutional review boards at universities and research institutions would do well to consider the implications when reviewing study proposals that involve the use of high-stakes-testing data. Individual scholars ought to consider them as well. Lately, I have been doing just that.

Since becoming a university faculty member in 2006, after many years as a high school teacher and school leader, I have been struck by the volume of research in education that relies on the results of high-stakes testing. Sometimes, this research is by scholars who are themselves critical of such testing. I myself feel a growing tension in my work between the practical need for objective indicators of student learning and the ethical demands of my profession to *do no harm*. So I have made a decision. From this point forward, to the best of my ability, I will not use the results of high-stakes tests as the sole or primary measure of student learning in research I conduct, unless a purpose of that work is to study the test itself.

As a junior faculty member who will be up for tenure in a few years, I realize this may complicate my future. But not doing so would complicate my conscience. I also recognize that the risks I face as an academic in standing up to the tests are substantially less than those my K-12 colleagues would face. The precollegiate reality is that funding for programs of direct benefit to children can be swiftly cut for acts of disobedience.

Still, I invite readers whose work involves education research to take the same pledge I have. Imagine the positive impact we could have on educational practice if we—college educators in particular—boycotted high-stakes testing. Whenever our work called for some yardstick of student learning, we would look somewhere *other than* the easily available and potentially harmful high-stakes tests.

Such an exercise would not only push our own practice constructively forward, but it might also lead to new ways of thinking about student assessment in general. And this new thinking could ripple out of academia and into pre-K-12 practice. The possibilities are exciting. The alternative is perilous. The stakes are high. Will you join me?

James H. Nehring, a former high school teacher, is now an assistant professor of education at the University of Massachusetts Lowell. He is the author of *The Practice of School Reform: Lessons from Two Centuries*, forthcoming from SUNY Press. He can be reached at james_nehring@uml.edu.