

## Are Educational Tests Inherently Evil?

Stephen G. Sireci

University of Massachusetts

Presidential Column for February 2007 Issue of the *NERA Researcher*  
(see <http://www.nera-education.org/>)

Tests are given for many reasons in the educational system. Many of these reasons are hated. For example, tests are major components in accountability systems that may have undesirable consequences for teachers, schools, or districts. Tests are also sometimes used as a requirement for something, such as high school graduation, scholarship, or eligibility to participate in collegiate athletics. In these instances, tests are often seen as a hurdle to overcome or as an unnecessary roadblock to an inherent right. Tests are also commonly used to assign grades to students, particularly beyond elementary school. Given these purposes, who could possibly like tests? The answer is hardly anyone, perhaps only the relatively few high achievers who enjoy a challenge or the opportunity to show what they (we?) can do.

Why are tests so widespread if they are so hated? Is it the same reason we have intelligent design and global warming? Of course not. The reason is that educational tests, if developed carefully, used properly, and interpreted appropriately, have enormous utility. As soon as all sides of the educational community acknowledge that fact, we can make progress toward a common goal of using assessments to improve student learning.

To properly understand educational tests, particularly their benefits and limitations, we must consider their use in specific situations. In this column, I discuss common perceptions and misconceptions of educational tests and the role tests currently play in federal and state education reform efforts. A primary goal of this discussion is to bridge the gap between proponents and opponents of standardized testing so that we can work together to improve student learning.

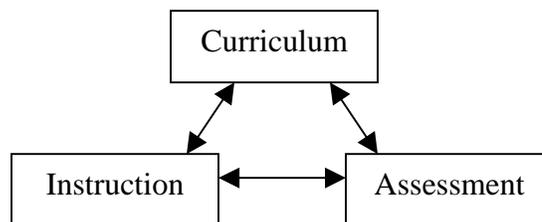
### *Why Are Tests So Ubiquitous in Education?*

A popular, but incorrect, myth is that educational are pushed by the extreme right of the political spectrum. This perception is simply false. Although the No Child Left Behind Act (NCLB) was proposed and signed by “he-who-must-not-be-named,” it was really an extension of Clinton’s *Goals 2000: Educate America* legislation, which was an extension of he-who-must-not-be-named’s father’s *America 2000* legislation. Thus, educational reform and accountability movements involving testing are one of the few bipartisan areas of legislation we have seen over the past several decades. There are, of course, strong differences in educational policy between Democrats and Republicans, such as the *financing* of education, but it is important to note that the NCLB Act was sponsored by Democrat Ted Kennedy and Republican Judd Gregg in the Senate and Democrat George Miller and Republican John Boehner in the House. It passed overwhelmingly in both (87-10 in the Senate and 381-41 in the House).

Why do federal legislators agree that mandated testing is an important part of education reform? There are several reasons. First, assessment is seen as a critical component in the educational process. In fact, quality education requires continuous interaction among instruction, curriculum, and assessment. Good instruction starts with good curricula and both influence each other. The development of curricula at the district and state levels is certainly influenced by what teachers teach in their classrooms. As the curricula are developed, teaching practices change accordingly. Assessments are needed to discover what students are learning. Based on that information, changes to instruction and curricula occur. My colleague Ron Hambleton refers to this dynamic interaction as the curriculum-assessment-instruction cycle, which is displayed in Figure 1. Alignment of these three components of the educational process is necessary for quality instruction.

Figure 1

The Curriculum-Instruction-Assessment Cycle



A second reason tests play a prominent role in federal and state education reform movements is that they are an effective means for quickly changing instructional practices. As McDonnell (2004) described “although standardized tests are primarily measurement tools to obtain information about student and school performance, they are also strategies for pursuing a variety of political goals” (p. 2). McDonnell also points out that there are few alternatives available to policy makers to enforce their educational policies. As she put it “Testing’s strong appeal is largely attributable to the lack of alternative policy strategies that fit the unique circumstances of public schooling...Standardized tests are one of the few, albeit incomplete, ways to measure outcomes of teaching” (p. 9).

A third reason mandated testing is a key component of education reform is that it forces educators to align their instruction with state curriculum frameworks. No teacher likes to be overly constrained regarding what she or he should teach. However, no one wants teachers spending large amounts of instructional time teaching knowledge and skills that most would consider unimportant, relative to other skills. Thus, education involves consensus about what should be taught. The development of curriculum frameworks without a means for assessing how well students master the objectives within them would create a situation in which the good work done in developing the frameworks could be simply ignored.

Critics of state-mandated testing argue that these tests narrow the curriculum and force teaching-to-the test. Proponents counter that the tests are aligned with curriculum frameworks, which were developed through a consensus process, and so teaching to the test is teaching to the frameworks. As in most debates, the truth probably lies somewhere in the middle. Nevertheless,

it is important to bear in mind that the idea behind consensus statewide curriculum frameworks and tests designed to measure them is a noble one, because its goal is to improve instruction. As a parent, I can understand this position. After all, I want to know that my sons' teachers are teaching the important knowledge and skills they will need to succeed personally and academically. State curriculum frameworks and tests designed to measure them attempt to ensure that what is taught is important. Thus, they aim toward facilitating quality instruction, as depicted in Figure 1.

### *Focusing on Test Use*

One of the greatest challenges we experience as educational researchers is asking the right research questions. With respect to educational testing, the questions we ask should be specific to using a test for a particular purpose. Thus, questions motivating research in this area should not be "Is the test bad?" or "Is the test fair?" but rather "Will the results of this test provide the information it is designed to produce?" Thus, evaluating a test means evaluating the use of a test for a particular purpose. Tests are not inherently "good" or inherently "bad," but using a test for some purpose could be either, depending on what the test was designed to do versus what it is used for. This notion is clear in the definition of *validity* presented in the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999):

Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. (p. 9)

As is evident in this definition, it is not a test that is validated per se, but the use of a test for a particular purpose. Defending inferences derived from test scores involves both qualitative evidence based on theories of what is being measured and quantitative evidence indicating the scores reflect the measured attribute. Thus, all educational researchers can contribute to research on testing, regardless of their particular research orientation. I will not discuss specific means for validating inferences in this column (see Kane, 1992 or Sireci, 2005 for examples). Instead, I focus on test use in educational testing and how it can help or hurt the educational process.

How are tests used in education? Teachers use tests to measure how well students grasp the material taught (e.g., classroom tests). Counselors use tests to diagnose students' strengths and weaknesses and make referrals for remediation, advanced courses, or other placement decisions. Policy makers use tests to evaluate teachers, schools, districts, states, countries, and various educational programs. Tests are also used as one criterion for high school graduation and for other types of certification such as an honors diploma, and for admissions into postsecondary and graduate education. As the stakes associated with educational tests increase, such as in the cases of granting a high school diploma or evaluating the performance of particular teachers, the criticisms also increase. And they should increase. If a test is used to make a "big" decision, the use of the test for that purpose should be supported by "big" evidence. Thus, as educational researchers, our assessment research activities should be focused on asking the right questions about test use (e.g., Is there evidence to support use of this test as a high school graduation requirement?).

I am a psychometrician working in educational measurement and so it is pretty obvious that I must believe in the usefulness of educational tests. However, my strong belief in the utility of educational tests stems not from my psychometric training, but from my experience as a parent. How do I know if my sons are receiving a good education? The class work, assignments, and report cards that come home give me some indication, but the norm-referenced and criterion-referenced test score reports give me a lot more to go on. The Iowa Tests of Basic Skills that our local school district uses allows me to compare my sons' performance to national norms. The Massachusetts Comprehensive Assessment System (MCAS) tests allow me to see how my sons are doing with respect to the performance standards established by the State. Now, when my wife and I speak with their teachers or the Principal, we can talk about these independent assessments, and how this information can be used to improve their instruction.

### *Looking Forward: Collaborating on Educational Assessment Research*

In this column, I merely touched on a few of the important issues that concern educational assessment policy and the proper use of tests in our nation's schools. I know there are many people who will never acknowledge the utility of a standardized test, but I also know there are many more who hate something else—bad teaching! Teachers who are not helping our students reach their academic potential are much more dangerous to our children than any test. I do not advocate using tests to “police” teachers or using test results to provide sanctions and rewards for teachers (a very bad idea, given the different types of students taught by different teachers). However, I like the idea of measuring students' achievement with respect to standards developed through a consensus process, and I like the idea of providing as much information as possible to parents and others about the academic achievement and progress of their children.

Are there problems with our current educational assessment policies? I think so. There are several valid criticisms about current educational tests. Personally, I am concerned about the amount our students are tested, and I am very concerned about the pressure that is put on students before they take a test. So, there is much room for improvement, which is where we, as educational researchers, come in. Let us not throw the baby out with the bathwater and simply dismiss tests as useless. Instead, let us research what seems to be working, what seems to be harmful, and what needs to be improved. By working together, we can improve educational assessment and provide advice to educational policy makers that is based on solid research. If we can do that, we will improve curriculum development, instruction, and assessment, with the happy consequence of improving student learning.

### *Closing Remarks*

I hope I have inspired everyone who read this column to focus on test use when thinking about educational assessments. Some of you may vehemently disagree with one or more of the points I raised. Others may agree, but have additional issues to bring to the discussion. In either case, I encourage you to write a short response for a future edition of the *NERA Researcher*, or write to me directly at [Sireci@acad.umass.edu](mailto:Sireci@acad.umass.edu). I seriously believe that everyone on both sides of the testing debate needs to work together to improve the instruction and learning experiences of our children. Please also contact me if there is an assessment issue you would like me to address

in a future edition of this column. And one other thing—don't forget to get your proposals in for NERA 2007!

#### References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.

Kane, M.T. (1992). An argument-based approach to validity. *Psychological Bulletin*, *112*, 527-535.

McDonnell, L. M. (2004). *Politics, persuasion, and educational testing*. Cambridge, MA: Harvard University Press.

Sireci, S. G. (2005). Validity theory and applications. *Encyclopedia of statistics in the behavioral sciences* (Volume 4, pp. 2103-2107). West Sussex, UK: John Wiley & Sons.