

## HIGH-STAKES TESTS: HARMFUL, PERMANENT, FIXABLE\*

W. James Popham

University of California, Los Angeles

The high-stakes tests currently used throughout American education are harming thousands of children—seriously and, perhaps, irreparably. But these tests, having been imposed because of a widespread belief that America’s schools weren’t working well enough, will not disappear in the foreseeable future. Accordingly, our nation’s high-stakes tests must be substantially improved so they supply accurate evidence for evaluating the nation’s schools, yet also support the instruction that goes on in those schools.

In quasi-syllogistic form, then, my three-point argument regarding the nation’s high-stakes tests runs as follows:

- *Current high-stakes tests are harming kids.*
- *High-stakes tests won’t go away.*
- *Therefore, high-stakes tests must be improved.*

*High-stakes tests defined.* I plan to deal, in turn, with each of these three points.

But first, one dollop of term-defining is requisite. A “high-stakes test,” to some, is a test

---

\*Adapted from a presentation at the annual meeting of the American Educational Research Association, April 1-5, 2002, New Orleans, Louisiana.

that, depending on a student's performance, is linked to important decisions about that student. Examples of such decisions would be the awarding of high-school diplomas, grade-to-grade promotion, or the requirement for low-scoring students to take remedial summer-school classes. Although I concur that such assessments are properly labeled as high-stakes tests, I also think the tests that are used to evaluate educators are high-stakes in nature. If school staffs, depending on students' test scores, are rewarded or sanctioned, then those tests are surely consequential to the educators involved. And even if no sticks or no carrots are involved, for example, if schools or districts are only ranked in local newspapers, I still regard those assessments as high-stakes. Newspaper rankings typically influence what goes on in schools. My definition of a high-stakes test, then, is any test for which students' performances have a meaningful impact either on those students or on the educators who taught them.

*Key components of accountability systems.* High-stakes tests have usually been installed in the U.S. as pivotal parts of *educational accountability* systems. These accountability mechanisms were typically established because of doubts on the part of our citizens that the schools were doing a good enough job.

The creation of accountability tests, of course, makes a good deal of sense. Educational policymakers, mirroring public skepticism about educational quality, wanted to devise evidence-producing procedures that could indicate whether our students were learning what they were supposed to be learning. If test scores were high, or were improving, then schools were thought to be doing a good job. If test scores were low, or

were not improving, an opposite conclusion was reached. Thus, with evidence of school-success at hand, policymakers believed they could more judiciously install programs designed to improve schooling where it needed to be improved. The only problem with this test-based formula for educational improvement was that *its architects used the wrong accountability tests!*

I am convinced that, with few exceptions, the educational policymakers who established America's test-based accountability systems were well-intentioned. They wanted our schools to improve. They wanted children to be well-taught. Unfortunately, these policymakers were appallingly ignorant of what kinds of tests should serve as the cornerstones of their accountability programs. Assessment illiteracy for an educational policymaker these days is a shortcoming far more insidious than any computer virus.

### **Harmful High-Stakes Tests**

Today's high-stakes tests are educationally injurious to an enormous number of American school children. I am convinced that this is so from having talked to hundreds of the nation's teachers in recent years. These teachers care about their students. Yet, they are saddled with pressures to boost their students' scores on the wrong kinds of high-stakes tests. And those pressures are real, not imagined. Those pressures are intense, not trivial.

Unfortunately, those score-boosting pressures are based on the wrong kinds of high-stakes tests. As a result, those pressures have led far too many teachers to engage in improper instructional practices. Let's look briefly at three of the educationally harmful consequences of teachers' attempts to raise students' scores on today's high-stakes tests.

*Curricular reductionism.* Students are being curricularly short-changed because teachers are being so relentlessly urged to raise students' test scores that many teachers give little or no attention to curricular content not apt to be assessed on their locally relevant high-stakes test.

This form of curricular tunnel-vision leads to reduced instructional attention for any subject falling outside the anointed realms of literacy and numeracy (and, in a few instances, science). Curricular reductionism also occurs *within* such fields as language arts or mathematics. Skills and knowledge in those subjects that, only a few years ago were regarded by teachers as indispensable, have been jettisoned if such content doesn't seem likely to be included on an applicable high-stakes test.

I have listened to more than a few teachers describe frightening instructional scenarios in which, for six weeks or more immediately prior to the administration of a statewide high-stakes test, *all instruction unrelated to the high-stakes test's content must cease*. These teachers had been directed by their principals or, possibly, by their district superintendents to "raise test scores at all costs." And that's just the problem,

it's the costs to kids of such test-worship. A curricularly short-changed child is a miseducated child. Today's high-stakes tests are leading to rampant curricular reductionism that, without debate, harms children.

*Excessive test-focused drilling.* Many teachers today are so intimidated by the need to raise students' scores on high-stakes tests that they transform their classrooms into drill-dominated training camps in which students are relentlessly drilled on tests or test-like exercises. Most teachers these days are familiar with the research showing that "time-on-task" is one of the most important variables influencing students' mastery. Therefore, because of the need to raise students' test scores, some teachers have installed so much task-time that their students' positive affect toward learning has surely been diminished.

Rather than deriving joy from the learning process, students in drudgery-drill classes can't wait for the period to end. The affective consequences for students of excessive test-focused drilling all too predictable. Those consequences are uniformly negative.

*Modeling dishonesty.* The third harmful impact of using the wrong kinds of high-stakes tests is that a number of pressured teachers have adopted test-preparation or test-administration practices that are blatantly dishonest. I don't approve of such conduct by any teacher, but I certainly understand why some teachers are driven by today's high-stakes tests to behave in ways that are clearly reprehensible.

Let's look at one common instance of such dishonesty. Some teachers have photocopied an *operational* form of a test, then given their students ample practice, with feedback, in answering slightly modified test items or, sometimes, in answering items that haven't been modified at all. When students subsequently, during the administration of the actual test, encounter a collection of previously practiced items, they immediately recognize that they have been made involuntary conspirators in a teacher-engineered fraud. If teachers are supposed to function *in loco parentis*, do we really want these proxy parents to serve as models of blatant dishonesty?

To repeat, I can see why teachers might succumb to such deplorable test-preparation practices. You see, in most states the operative high-stakes tests simply do not set forth *with sufficient clarity for day-to-day instructional decision-making* just what skills and knowledge are to be assessed. The besieged teacher is directed to raise test scores, but can't get an adequate fix on just what is supposed to be tested.

And that's true even in states where customized "standards-based" tests have been created to measure the state's officially approved content standards (knowledge and skills). Typically, there are so many content standards to be taught that teachers really can't figure out what content will actually be measured by each year's high-stakes tests. Moreover, after the high-stakes tests have been administered, no standard-by-standard reporting is provided. So, neither teachers nor students discover which content standards have or haven't been mastered. Is it any wonder that some

frustrated teachers stoop to unethical test preparation or improper test administration such as supplying excessive “help” to test-takers?

But understanding the causes for unprofessional conduct by classroom teachers does not excuse those teachers. Teacher dishonesty tends to breed dishonesty from students. Cheating in our classrooms is inexcusable—on either side of the teacher’s desk.

So, summing up, today’s high-stakes tests have fostered curricular reductionism, excessive test-focused drilling, and dishonest conduct by teachers. How widespread are these three consequences of high-stakes tests? Well, I believe that there is much more of this educational mischief going out there than anyone wants to admit. Of course, if any of these negative things were happening in a *single* classroom, that would be one classroom too many.

### **Permanent High-Stakes Tests**

American education’s high-stakes tests are here to stay. They were installed by policymakers who doubted that the nation’s educators were doing a good job.

Unfortunately, there has been little evidence to disabuse educational policymakers of those doubts. Indeed, the push for more rigorous educational accountability has been an almost automatic political plank in the platforms of those who now seek office at the state or federal levels.

There was a time, perhaps a half-century ago, when the public would accept, and actually believe, proclamations from educational administrators that our schools were in good shape. But during those 50 years we have seen far too many illiterate students snaring diplomas on the basis of seat-time. We have seen far too many disquieting comparisons between our students' test scores and those of students elsewhere in the world. We have simply seen far too many low test scores.

It makes little difference that some of these low test performances may have taken place on the wrong tests. Low test scores, to the U.S. public, translate into a perception of unsuccessful schools. And unsuccessful schools, according to those who set educational policies, definitely need to be *monitored*.

Accountability programs, built upon accountability tests, have become a key component of America's educational Zeitgeist. High-stakes tests are not going to go away, at least in our lifetime. Those who call for the abolition of high-stakes tests, therefore, are spitting into the wind.

### **Fixable High-Stakes Tests**

Because correct high-stakes tests are harming students, and because those tests are not going to vanish, we need to fix them. And we can.

The problem with today's high-stakes tests, in an oversimplified nutshell, is that those tests have been constructed according to a traditional measurement paradigm aimed at providing comparative interpretations of test-takers' performances. And, in order to create the amount of score-spread necessary to yield meaningful comparisons based on the short amount of time available for such testing, a traditional test's items need to do a potent job in spreading out students' scores. Many of the items in today's high-stakes tests, because of the relentless need for score-spread, are linked directly to students' socioeconomic status (SES) or to students' inherited academic aptitudes (such as children's inborn verbal, quantitative, and spatial potentials). SES-linked items and inheritance-linked items do a good job in spreading out students' scores. But these sorts of items *measure what students bring to school, not what they learn there.*

Tests that include many items measuring what students bring to school, rather than what students learn in school, are obviously inappropriate for evaluating a school's instructional success. To measure a school's quality with today's high-stakes tests is like trying to determine the winner of a NASCAR race by a pre-start appraisal of what's in each car's fuel tank.

Moreover, as noted earlier, current high-stakes tests (1) fail to provide educators with clear descriptions of the content standards to be measured and (2) do not yield the kinds of standard-by-standard results that would allow teachers to improve their instruction.

But it *is* possible to create accurate accountability tests, yet do so in a manner that enhances the quality of teachers' instructional decisions. At present, the best guidance regarding how to create more defensible educational accountability tests is represented in the nine requirements set forth by the Commission on Instructionally Supportive Assessment. That independent commission, convened in mid-2001 by five national organizations of educators, provided two reports in October 2001.\* Both reports, one for policymakers and one for the individuals who must create high-stakes accountability tests, are available online from any of the five convening associations.\*\*

I agree totally with the commission's prescription about how to improve high-stakes tests so they (1) help rather than harm children's education and (2) yield accurate evidence by which to evaluate schools. Accordingly, listed below are the unadorned nine commission requirements. Anyone interested in the rationales underlying the nine requirements, or in procedures that might be employed to satisfy the nine requirements, should consult the commission's two reports.

### The Nine Requirements of the Commission on Instructionally Supportive Assessment

---

\*The Commission on Instructionally Supportive Assessment. (1) *Building Tests That Support Instruction and Accountability: A Guide for Policymakers*, (2) *Illustrative Language for an RFP to Build Tests That Support Instruction and Accountability*. Washington, DC: Author, 2001.

\*\* The Commission on Instructionally Supportive Assessment was convened by the American Association of School Administrators, the National Association of Elementary School Principals, the National Association of Secondary School Principals, the National Education Association, and the National Middle School Association. Available online at: [www.aasa.org](http://www.aasa.org), [www.naesp.org](http://www.naesp.org), [www.principals.org](http://www.principals.org), [www.nea.org](http://www.nea.org), and [www.nmsa.org](http://www.nmsa.org).

*Requirement 1. Prioritized Content Standards:* A state's content standards must be prioritized to support effective instruction and assessment.

*Requirement 2. Unambiguously Described Content Standards:*

A state's high-priority content standards must be clearly and thoroughly described so that the knowledge and skills students need to demonstrate competence are evident.

*Requirement 3. Standard-by-Standard Reporting:* The results of a state's assessment of high-priority content standards should be reported standard-by-standard for each student, school, and district.

*Requirement 4. Classroom Assessments for State-Unassessed Content Standards:* A state must provide educators with optional classroom assessment procedures that can measure students' progress in attaining content standards not assessed by state tests.

*Requirement 5. Monitoring Curricular Breadth:* A state must monitor the breadth of the curriculum to ensure that instructional attention is given to all content standards and subject areas, including those that are not assessed by state tests.

*Requirement 6. Appropriate Assessment for All Students:* A state must ensure that all students have the opportunity to demonstrate their achievement of state standards; consequently, it must provide well-designed assessments appropriate for a broad range of students, with accommodations and alternate methods of assessment available for students who need them.

*Requirement 7. Sufficient Test-Development Time:* A state must generally allow test developers a minimum of three years to produce statewide tests that satisfy the Standards for Educational and Psychological Testing and similar test-quality guidelines.\*

*Requirement 8. Pertinent Professional Development:* A state must ensure that educators receive professional development focused on how to optimize children's learning based on the results of instructionally supportive assessments.

*Requirement 9. Ongoing Evaluation:* A state should secure evidence that supports the ongoing improvement of its state assessments to ensure those assessments are (a) appropriate for the accountability purposes for which they are used, (b) appropriate for determining whether students

---

\* See American Educational Research Association. 1999. *Standards for Educational and Psychological Testing*, Washington, DC: Author. See also such documents as Pellegrino, J.W., N. Chudowsky, and R. Glaser (Eds.) 2001. *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academy Press.

have attained state standards, (c) appropriate for enhancing instruction, and (d) not the cause of negative consequences.

### **An Alternative**

I think what's going on in American schools today as a direct consequence of high-stakes testing is deplorable. The education being received by growing numbers of our children is deteriorating because the wrong kinds of high-stakes tests are being used.

But, from a practical and political perspective, high-stakes tests will not disappear from our nation's schools. Too many citizens doubt that American schools are doing a satisfactory job of educating the nation's youth. Quite reasonably, policymakers want evidence to indicate which schools are successful and which schools aren't. High-stakes accountability tests, therefore, will be with us for a long while.

It is possible, however, to install high-stakes tests that satisfy the need for accountability evidence, yet also stimulate improved instruction. If high-stakes tests were created in accord with the October 2001 recommendations of the Commission on Instructionally Supportive Assessment, the negative impact of American education's high-stakes testing could be halted. It should be.