

First, Do No Harm

Although high standards hold great promise for improving teaching and learning for low-achieving students, the improper use of graduation and promotion tests can hurt the very students who need help the most.

Jay P. Heubert

The central and oft-stated goal of the standards and accountability movement is to help all students achieve high levels of knowledge and skill. Certainly no one needs high-quality curriculum and teaching more than low-achieving students, who have long been the victims of low expectations, weak instruction, and inadequately funded schools. Data from large-scale assessments, if used properly with other relevant information, can help improve curriculum and pedagogy, hold schools accountable for improved achievement, and identify and address students' learning needs. And pass rates on state graduation tests appear to be rising, at least before such factors as dropout rates and increased retention in grade are taken into account.

Even so, in more and more states and school districts, testing programs and education practices that are supposed to help low-achieving students are instead putting those students at substantially increased risk of leaving school without the standard high school diplomas that open the doors to much that is good in life. In our society, not having a diploma is associated with lower pay, diminished opportunities for

employment and further education, higher risk of criminal incarceration, and greater likelihood of dysfunction in family life (National Research Council [NRC], 1999). Moreover, "alternative" diplomas are no match for the real thing; evidence suggests that holders of the General Equivalency Diploma resemble high school dropouts in terms of future educational and employment opportunities (NRC, 2001).

We should therefore be concerned, to put it mildly, that students of color, students with disabilities, English-language learners, and low-income students are failing some demanding state graduation tests at rates as high as 60 to 90 percent. And these failure rates would be even higher if they included the many students who drop out or are retained in grade before they even take graduation exams (Heubert, in press).

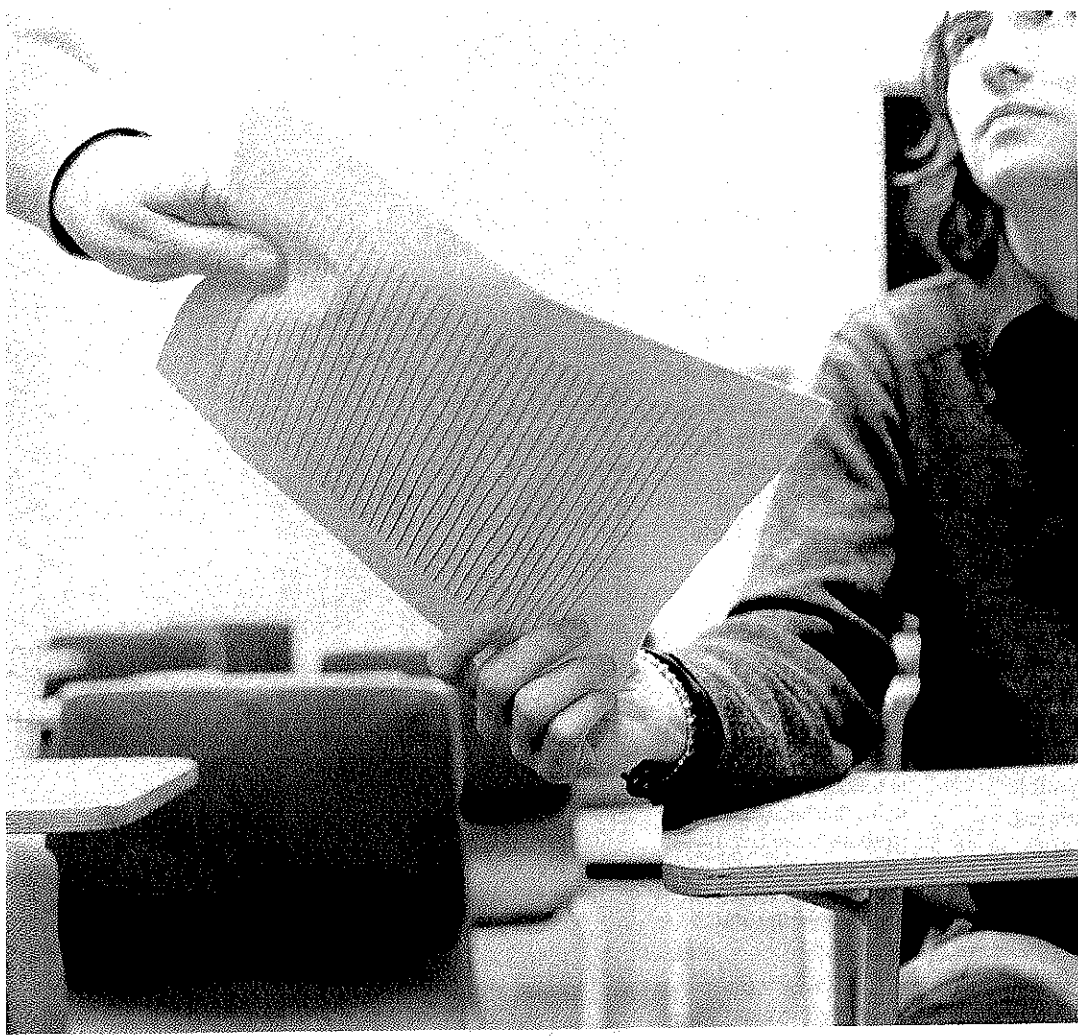
Indeed, pass-rate gaps by race, ethnicity, language proficiency, disability, and income have actually widened in some states—even before accounting for dropouts and retention. The racial achievement gap among 12th graders in the United States has been increasing for some time, even during a decade in which states have been adopting higher standards and have

begun to make changes in teaching and learning aimed at helping all students meet those standards.

Are there conditions under which high standards and demanding assessments, intended above all to help low achievers, do more harm than good for the students who most need our help? Research indicates that these conditions do exist and suggests that many states and school districts use graduation and promotion tests improperly, in ways that diminish rather than enhance the life chances of many low-achieving students.

The good news is that if states and school districts refrained from improper use of promotion and graduation tests—like physicians who pledge "first, do no harm"—high standards and challenging assessments would be far likelier to help rather than hurt our neediest students.

Among the policies that educators and policymakers should avoid in order to refrain from doing harm, three stand out: retaining students in grade, placing students in low-track classes, and requiring students to pass graduation or promotion tests before schools have given the students an opportunity to acquire the knowledge and skills that such tests measure.



© David Paul Productions/Getty Images

tion tests has remained stable at around 18, the number of state promotion-test programs nearly tripled, from 6 to 17, just between 1999 and 2001. Moreover, many urban school districts—Chicago, Boston, and New York City, for example—have adopted promotion-test policies even when their states have not. And further proliferation is likely: Although federal law neither requires nor forbids promotion testing, some states already plan to use for promotion purposes the new tests that will soon be required in grades 3 through 8 by the No Child Left Behind Act (2002).

Thus, in the United States, most minority and immigrant students—and increasing numbers of all students—are subject to state or local promotion-test programs, placing them at increased risk of experiencing the serious harms associated with retention and dropping out. We know better.

Retention in Grade

Decades of high-quality research show that students required to repeat a grade are much worse off than similar students who are instead promoted to the next grade. The National Research Council (1999) found that low-performing elementary and secondary school students who are held back—as compared with equally weak students who are promoted—do less well academically, are much worse off socially, and are far likelier to drop out. Recent studies (Goldschmidt & Wang, 1999; Lillard & DeCicca, 2001) conclude that retention in grade is the single strongest predictor of which students will drop out—stronger even than parental income or mother's education level. The evidence on simple retention in grade was so compelling

that the National Research Council (1999) said it constituted inappropriate test use for states or school districts to use test scores for this purpose.

Despite the common-sense notion that promoting students who do not seem ready for the next grade is unwise, the research shows that holding students back is even worse. Better than either retention or social promotion, of course, is to identify and address student needs early, before students fail a promotion exam. Many kindergarten and 1st grade teachers can reliably predict which of their students will probably experience later difficulties with 3rd grade promotion tests.

And yet, bipartisan calls to “end social promotion” have helped produce dramatic increases in promotion testing. Even as the number of states with gradua-

Assignment to Low-Track Classes

As with retention, decades of good research have revealed the harmful educational consequences of placing students in low-track classes. According to the National Research Council,

Low-track classes are typically characterized by an exclusive focus on basic skills, low expectations, and the least-qualified teachers. Students assigned to low-track classes are worse off than they would be in other placements. This form of tracking should be eliminated. Neither test scores nor other information should be used to place students in such classes. (1999, pp. 280-281)

The standards movement rests on the premise that virtually all students can

reach high levels of achievement if they receive high-quality curriculum and instruction. This premise rests, in turn, on dramatic recent research findings in such areas as brain development, early childhood education, and effective pedagogy. In at least three statutes, the U.S. Congress has accepted this premise and the research supporting it. Since 1994, federal law has specifically required that low-achieving disadvantaged students receive "accelerated," "enriched," and "high-quality" curriculums, "effective instructional strategies," and "highly qualified instructional staff" (Improving America's Schools Act, 1994).

Low-track classes obviously reflect very different assumptions: that many students can't learn beyond low levels, that good teachers are wasted on low achievers, and that high expectations will destroy the self-esteem of low achieving students.

These assumptions are incompatible with the modern standards movement, with federal law, and with efforts to help all students attain high-level knowledge and skills. And yet many states, school districts, and schools assign students to dead-end, low-track classes in which they fall further behind each year instead of acquiring increasingly high-level knowledge and skills. We know better.

Holding Students Accountable for Subject Matter Not Yet Taught

Educators and experts in testing and measurement agree that student test scores should not be used to make individual promotion and graduation decisions until after students have had an adequate opportunity to acquire the knowledge and skills that such tests measure. In short, it is improper to punish individual students for not knowing what their schools have never taught them. Court cases involving graduation tests reach similar conclusions.

For example, the testing profession's bible—*Standards for Educational and Psychological Testing*, issued jointly by

the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education—states unequivocally that promotion and graduation tests should cover *only* the "content and skills that students have had an opportunity to learn" (1999, p. 146). Similarly, when the U.S. Congress asked the National Academy of Sciences to study high-stakes testing, the resulting report concluded that

tests should be used for high-stakes decisions . . . only after schools have implemented changes in teaching and curriculum that ensure that students have been taught the knowledge and skills on which they will be tested. (NRC, 1999, p. 6)

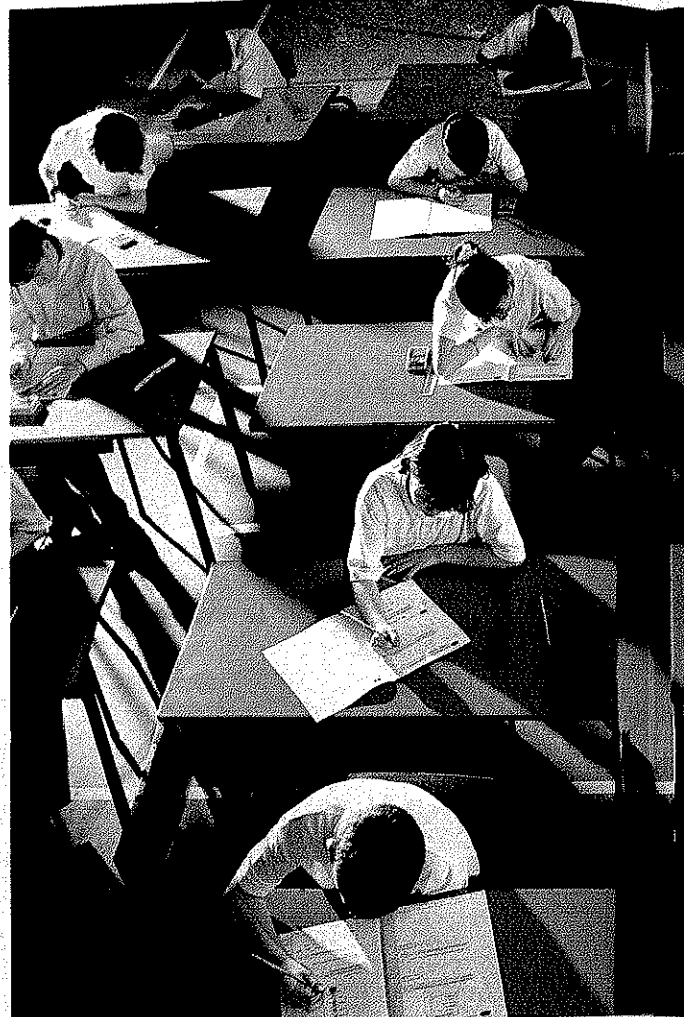
Courts have ruled for two decades that graduation tests must measure "that which was taught" (*Debra P. v. Burlington*, 1981).

How can educators, policymakers, and parents determine whether schools are teaching students the knowledge and skills that tests measure? They can begin by checking three important kinds of data:

Check the failure rates. Perhaps the most straightforward approach is to examine actual indicators of student achievement, such as test scores and grades. According to one recent study,

the best evidence that a school system is providing its students adequate opportunity to learn the required material is whether most students do, in fact, learn the material. (Wise et al., 2002, p. 93)

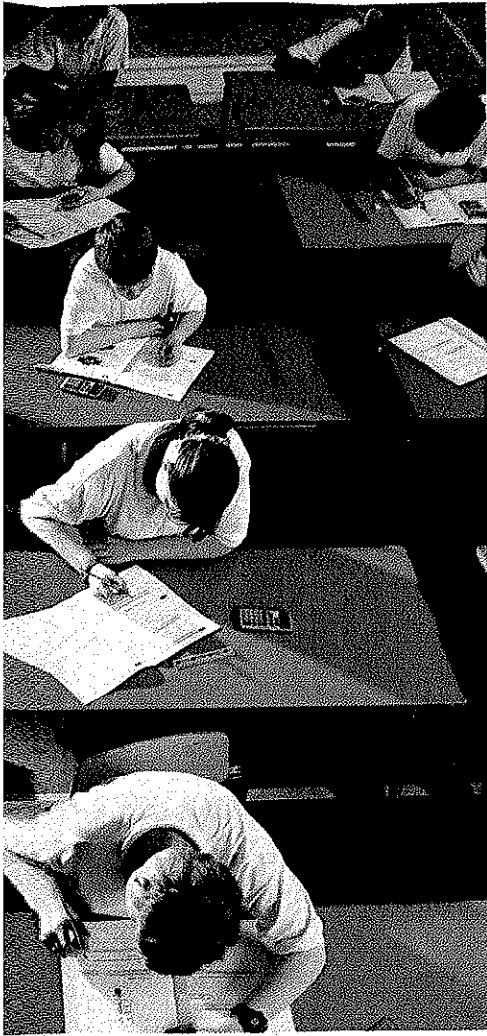
By this measure, very high failure



rates on many promotion and graduation tests raise real doubt that students have received a reasonable opportunity to acquire the requisite knowledge and skill. If virtually all students can learn to high standards, such high failure rates must be due, at least in part, to a shortage of high-quality instruction. The fact that schools that serve many low-achieving students often have the lowest funding levels and the lowest proportions of well-qualified teachers supports this conclusion.

Check the disaggregated data. Although federal law does not require states or school districts to administer tests that have high stakes for individual students, it does require that states and school districts

■ Set the same high standards for all students;



© Steve Lymeworkbook/Stock

- Include all students in statewide assessments—including, with appropriate accommodation, students with disabilities and English-language learners; and

- Report scores disaggregated by race, disability, English proficiency, and family income.

Although compliance with federal law is improving, many states do not yet include all students with disabilities or students with limited English proficiency in their assessment systems, and many do not yet disaggregate achievement information properly for various student populations (Robelin, 2001). Without such data, states and school districts lack basic information about how well low-achieving groups are performing and how to serve them more effectively. And if such informa-

tion is a precondition to improved school performance, the lack of data suggests that many states are not yet educating students well enough to hold them accountable on promotion and graduation tests.

Check the alignment. Are curriculum and instruction properly aligned with state standards and state assessments? Do schools and teachers possess the capacity to deliver high-quality instruction to all students? The logic of the standards movement is that improved alignment and capacity are the principal means to improved student achievement.

large numbers of low-achieving students, problems aligning instruction for students with disabilities with state standards and tests, and budget cuts in the state programs designed to help low-achievers reach high standards. In sum, research suggests that many states and schools are not yet at the point where they are teaching all students the high-level knowledge and skills that high-stakes tests increasingly measure (Heubert, in press).

Capacity and alignment should improve over time, but weak alignment between what teachers teach and what state tests measure presents a serious

Very high failure rates on many promotion and graduation tests raise real doubt that students have received a reasonable opportunity to acquire the requisite knowledge and skills.

Despite improvements, many schools have not yet sufficiently aligned their curriculum and teaching with state tests. For example, a recent study of 11 states (Porter & Smithson, 2000), including several with promotion or graduation tests, revealed only modest overlap between what a state's tests measure and what that state's teachers say they teach. For example, in 4th grade mathematics, teachers reported overlaps of 45 to 23 percent, which means that they were not teaching 55 to 77 percent of what their state test measured. In 8th grade mathematics, reports from six states showed even less overlap, ranging from a high of 35 percent to a low of 5 percent. Figures in 4th grade science were comparable, and those from 8th grade science showed even less overlap. A related multistate study conducted in 2001 concluded that "instructional content was not very well aligned with . . . the state test" (Blank, Porter, & Smithson, 2001, p. 26).

Other studies document major problems of capacity in schools that serve

problem when schools use tests to make high-stakes decisions about individual students. According to professional standards for testing and court decisions involving graduation testing, close alignment should precede the use of test scores for high-stakes purposes.

Unfortunately, current alignment and capacity problems have not prevented more states and school districts from adopting and implementing high-stakes testing requirements for individual students, or from raising the bar that students must clear to receive standard diplomas or advance to the next grade.

This sequence is wrong. It threatens to punish students whose only offense is that they do not know what their schools have not yet taught them. We know better.

We Can Do Better

Like physicians who pledge "first, do no harm," policymakers, educators, test developers and researchers can strive to avoid policies and practices that are inimical to the attainment of high educational standards for all.

Low-achieving students tend to rely heavily on their schools for academic knowledge and skills; they need high-quality instruction more than anyone else. That so many of these students have been ill-served by their schools for so long is a tragedy; that states and schools now aspire to educate all students to high standards is a welcome change.

documented harms associated with grade retention and denial of high school diplomas. Those at greatest risk include the very populations—students of color, students with disabilities, English-language learners, and low-income students—whom standards-based reform could potentially help the most. We can do better. ■

Without such data, states and school districts lack basic information about how well low-achieving groups are performing and how to serve them more effectively.

Test-score data can play an important diagnostic role. Data from good tests, used properly and in combination with information from other sources, can help identify strengths to build on and weaknesses to address: in states, school districts, schools, and classrooms; in curriculum materials and teaching techniques; and for individual educators and students. Test-score information can be a lever for improving teaching and learning; it should not be a means by which we punish students for not knowing what their schools have not yet taught them.

Unfortunately, the improper use of promotion and graduation tests threatens to distract us from our central objectives. Especially when the economy is weak and budgets are tight, policymakers may often find it quicker, cheaper, and more palatable politically to adopt test requirements that hold individual students accountable than to do the hard, time-consuming work of identifying effective education practices and making the many changes in curriculum and pedagogy that would help prepare all students to meet high standards.

In too many places, however, graduation and promotion tests are putting many students at sharply increased risk of suffering the serious, well-

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Blank, R., Porter, A., & Smithson, J. (2001). *New tools for analyzing teaching, curriculum, and standards in mathematics and science: Results from survey of enacted curriculum project final report*. Washington, DC: Council of Chief State School Officers.
- Debra P. v. Turlington*, 474 F. Supp. 244 (M.D. Fla. 1979); aff'd in part and rev'd in part, 644 F. 2d 397 (5th Cir. 1981); rem'd, 564 F. Supp. 177 (M.D. Fla. 1983); aff'd, 730 F. 2d 1405 (11th Cir. 1984).
- Goldschmidt, P., & Wang, J. (1999). When can schools affect dropout behavior? A longitudinal multilevel analysis. *American Educational Research Journal*, 36(5), 715-738.
- Heubert, J. P. (in press). High-stakes testing in a changing environment: Disparate impact, opportunity to learn, and current legal protections. In S. Fuhrman & R. Elmore (Eds.), *Redesigning accountability systems*. New York: Teachers College Press.
- Improving America's Schools Act, 20 U.S.C. sections 6301 et seq. (1994).
- Lillard, D., & DeCicca, P. (2001). Higher standards, more dropouts? Evidence within and across time. *Economics of Education Review*, 20(5), 459-474.
- National Research Council, J. P. Heubert, & R. Hauser (Eds.). (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press.
- National Research Council, A. Beatty, U. Neisser, W. Trent, & J. P. Heubert (Eds.). (2001). *Understanding dropouts: Statistics, strategies, and high-stakes testing*. Washington, DC: National Academy Press.
- No Child Left Behind Act of 2001, Public Law 107-110 (January 8, 2002).
- Porter, A., & Smithson, J. (2000, April). *Alignment of state testing programs, NAEP, and reports of teacher practice in grades 4 and 8*. Paper presented at the annual meeting of the American Educational Research Association. New Orleans, LA.
- Robelin, E. (2001). States sluggish on execution of 1994 ESEA. *Education Week*, 21(13), 1, 26, & 27.
- Wise, L., Sipes, D. E., Harris, C., George, C., Ford, J., & Sun, S. (2002). *Independent evaluation of the California High School Exit Examination (CAHSEE): Analysis of the 2001 administration*. Sacramento, CA: California Dept. of Education.

Author's note: This article was written with support from the Carnegie Scholars Program of the Carnegie Corporation of New York and from the National Center on Accessing the General Curriculum, pursuant to a cooperative agreement between the Center for Applied Special Technologies and the U.S. Department of Education, Office of Special Education Programs. The opinions expressed are the author's; they do not necessarily reflect the position of the sponsors, and no endorsement should be inferred.

For a fully documented analysis of related issues, please see my chapter, "High-Stakes Testing in a Changing Environment: Disparate Impact, Opportunity to Learn, and Current Legal Protections," in *Redesigning Accountability Systems*, edited by S. Fuhrman & R. Elmore (in press) from Teachers College Press.

Copyright © 2002 Jay P. Heubert.

Jay P. Heubert is an associate professor of education at Teachers College, Columbia University, 212E Main Hall, Box 157, 525 W. 120th St., New York, NY 10027; jay.heubert@columbia.edu. He is also an adjunct associate professor at Columbia Law School.