



Research Link

The Use and Misuse of Standardized Tests

John H. Holloway

In the United States, the use of nationally developed standardized tests has proliferated during the past decade (Council of Chief State School Officers, 2000). In a recent survey of U.S. public attitudes about education reform (Hart & Teeter, 2001), almost one-half of the respondents touted the benefits of these tests for improving education, but one-third expressed concerns about their possible misuse. The survey identified two recurrent themes: Testing is important in education reform, but we need to use tests carefully.

What Tests Can Measure

What purposes do tests like the National Assessment of Educational Progress (NAEP), Stanford 9, the Iowa Tests of Basic Skills, and the ACT and SAT serve?

The NAEP surveys the educational accomplishments of groups of U.S. students in a variety of subjects. It reports the results by student populations—grade levels, for example—and subgroups of those populations, such as male or Hispanic students, but does not provide individual scores of students or schools. The reporting metrics used by the NAEP allow performance comparisons within a subject from year to year and from one subgroup of students to another in the same grade (Brown, Dabbs, Kostad, & Horkay, 1999).

Unlike the NAEP, the Stanford 9 and the Iowa Tests of Basic Skills report individual student achievement. The questions on these tests align with standards that have been developed by such national organizations as the National Council of Teachers of Mathematics, the

American Association for the Advancement of Science, and the National Council for the Social Studies, among others (Harcourt, 2001; Riverside, 2001). The Stanford 9 combines open-ended and multiple-choice questions to measure students' academic achievement, and the Iowa Tests of Basic Skills provides a comprehensive assessment of student progress in basic skills, including the ability to read maps and locate and evaluate different sources of information.

Used for their intended purposes, assessments can assist education reform.

Aiming beyond the K-12 curriculum, the ACT Assessment and the College Board's SAT-I and SAT-II attempt to predict students' readiness to undertake college-level work. Most colleges consider these tests' results as part of evaluating applicants for admission. The ACT measures general educational development in English, reading, mathematics, and science (ACT, 2001), and the SAT-II assesses achievement in such subjects areas as English, biology, and foreign languages (College Board, 2001). The SAT-I, on the other hand, measures verbal and mathematical reasoning abilities that students develop both in and out of school; it does not test any particular state, school, or district curriculum or high school course (College Board, 2001).

Asking Too Much of Tests

Used for their intended purposes, these assessments can assist education reform by tracking the progress and levels of

achievement of individuals or groups of students and by indicating who is ready to tackle college-level work. Unfortunately, the pressure on educators and policymakers to demonstrate accountability in schools has driven some to use the test results inappropriately. The *NAEP Guide* cautions that

casual inferences related to subgroup membership, the effectiveness of public and nonpublic schools, and state or district-level educational systems cannot be drawn using NAEP results. (Brown, Dabbs, Kostad, & Horkay, 1999, p. 30)

Nonetheless, Donald Gratz (2000) points out that "educational accountability is still in its infancy" and "testing is often handled poorly" (p. 681).

To address the concerns raised by Gratz, David Pearson and his colleagues (2001) contend that we must not ask the tests to perform tasks that they cannot do. Karen Mitchell (1997) agrees that many reform efforts have been derailed by misaligned assessment, and she cautions principals and teachers to track performance on a broad range of outcomes over time.

Ensuring Credibility and Validity

Researcher Robert Linn (2000) finds that assessment in the reform movement continues to emphasize accountability but now also focuses on rigorous content, performance standards, and the inclusion of all students. Linn also notes that test results often connect to high-stakes accountability for schools, teachers, and students. Linn suggests ways to improve the credibility and validity of the assessment system and to overcome some of its negative effects:

- Provide safeguards against the exclusion of students from assessments.

■ Use multiple indicators instead of a single test.

■ Emphasize the comparison of performance from year to year rather than from school to school.

■ Consider value-added systems, which provide schools a reasonable chance to show improvement.

■ Recognize, evaluate, and report the degree of uncertainty in the results.

■ Evaluate both the intended positive effects and the more likely unintended negative effects of the system.

In addition to Linn's advice, the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (1999) have created standards for using tests and offer several suggestions for improving the trustworthiness and authority of assessment:

■ Monitor the consequences of tests and identify and minimize the potential for negative consequences.

■ Accompany reports of group differences in test scores with relevant contextual information. Caution users against misinterpretation.

■ Explain supplemental information to minimize possible misinterpretations of the data.

■ Ensure that the individuals who make decisions within the school or program are proficient in the appropriate methods for interpreting test results.

If educators and stakeholders have a thorough understanding of how to use and interpret assessment results, they will have powerful opportunities to bring about sustained, systemic improvement in our schools. ■

References

- ACT. (2001). *ACT Assessment* [Online]. Available: www.act.org/aap/index.html
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Brown, J., Dabbs, P., Kostad, A., & Horkay, N. (1999). *The NAEP guide*. Washington, DC: National Center for Education Statistics, Office of Educational Research and Improvement, U. S. Department of Education. Available: <http://nces.ed.gov/nationsreportcard/pdf/main1999/2000456.pdf>
- College Board. (2001). *How the SAT is made*. New York: The College Board. Available: www.collegeboard.org/html/pdf/howsatmade.pdf
- Council of Chief State School Officers. (2000). *Key state education policies on K-12 education: 2000*. Washington, DC: Author. Available: www.ccsso.org/pdfs/KeyState2000.pdf
- Gratz, D. B. (2000). High standards for whom? *Phi Delta Kappan*, 81, 681-687. Available: www.pdkintl.org/kappan/kgra0005.htm
- Harcourt. (2001). Stanford 9 content overview [Online]. Available: www.hemweb.com/trophy/achvtst/content.htm
- Hart, P. & Teeter, R. (2001, May). *A measured response: Americans speak on education reform*. Princeton, NJ: Educational Testing Service. Available: [ftp://etsi1.ets.org/pub/corp/2001_executive_report.pdf](http://etsi1.ets.org/pub/corp/2001_executive_report.pdf)
- Linn, R. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4-17. Available: www.aera.net/pubs/er/arts/29-02/linn01.htm
- Mitchell, K. (1997). What happens when school reform and accountability testing meet? *Theory into Practice*, 36(4), 262-265.
- Pearson, P. D., Vyas, S., Sensale, L., & Kim, Y. (2001). Making our way through the assessment and accountability maze: Where do we go now? *The Clearing House*, 74(4), 175-182.
- Riverside. (2001). Iowa tests of basic skills, form M [Online]. Available: www.riverpub.com/products/group/itbs_indepth.htm

John H. Holloway is Project Director, Educational Testing Service, Princeton, NJ 08541; jholloway@ets.org.

The Educator's Essential

Are you intrigued by brain research but unsure of what to do with it?

Announcing the release of

Exceeding Expectations: A User's Guide to Implementing Brain Research in the Classroom

..... -By Susan Kovalik & Karen D. Olsen

This exciting new book describes the brain research base for the Kovalik ITI model and provides hundreds of curriculum and instruction practices for acting on it.

ITI is used successfully in thousands of schools across the country. It is approved for Comprehensive School Reform Demonstration (CSRD) funding and listed in the *Catalog of School Reform Models*.

Read now to gain a coherent, understandable, up-to-the minute view of brain research in action - *theory to application*.

Toll free: 888-777-9827

Fax: 253-630-7215

E-mail: book4@oz.net
www.books4educ.com

Order Today



