

What Parents Should Know About Test Accuracy and Use

The accuracy and fairness of standardized testing is taken very seriously in the education world. These issues are a major focus of both the testing experts who develop standardized tests and the researchers who endeavor to ensure a test's fairness, reliability, validity, and accuracy. But many issues remain both controversial and complex. Here's a start in looking at these questions.

I've always thought that standardized test scores were accurate. But my son's teacher said they are just one indicator of how my son is doing in school. How accurate are test scores?

Good question! In general, tests are designed to provide dependable and valuable information about student achievement or aptitude. At best, they provide a source of objective information for decisions and judgments that otherwise might be subjective, arbitrary, or inconsistent. At the same time, many issues in testing about perceived economic, cultural, and gender bias continue to be raised. Expert educational test designers work hard to try to ensure that standardized tests accurately measure what they are designed to measure and are as objective and unbiased as possible. However, these are very complex tasks.

Your question about test accuracy raises at least three related issues. One concerns how well (how fully and accurately) a *single test* score can evaluate a person's knowledge and abilities. Another concerns the accuracy of test scores, including their scoring and reliability. And another focuses on the uses that are made of tests and test results, because a test that works well for one purpose may not provide accurate or reliable information when used for a different purpose.

No Single Test

The National Research Council (NRC) of the National Academy of Sciences issued a report in 1999 from its Board on Testing and Assessment. The report found that educational tests generally do provide dependable and valuable information about student achievement, but that they are definitely not perfect. The researchers pointed out "... a test score is not an exact measure of a student's knowledge or skills...no single test score can be considered a definitive measure of a student's knowledge." The report added that large-scale tests often use different versions of the test *form* to prevent cheating and that an

Resources

Helpful Web Sites

CAESL<http://www.caesl.org/>**CRESST**<http://www.cse.ucla.edu/>**WestEd**<http://www.wested.org/>**Parent Portal at LHS**<http://lhsparent.org>**Greatschools.net**<http://www.greatschools.net/>**National PTA**<http://www.pta.org/>**National Parent Information Network**<http://www.NPIN.org/>**Family Education Network**<http://www.familyeducation.com>

individual's score can be expected to vary somewhat across different forms of a test—even though test developers try to keep the forms at about the same level of difficulty. Scores can also vary due to “transitory factors,” such as the student’s health on the day of the test, test anxiety, and other testing conditions. Some research shows that a student taking the same test twice, even just a month apart, rarely scores the same each time, and sometimes considerably different. Student knowledge, skills, and performance vary from month-to-month and even from day-to-day, contributing to questions about which is the more accurate result.

Test Design and Scoring

Accuracy is an important consideration in many aspects of testing, including test design, measurement error, and scoring accuracy. Good test design is necessary in the development of a test that will serve its intended purpose well. There are many factors that go into determining reliability. This involves careful research with samples of the population to be tested to determine whether the testing instruments actually measure what they are designed to, as well as statistical analysis. Most large-scale testing is conducted by educational testing companies that have careful protocols and procedures. Mistakes in test scoring do not happen very often, but they can occur. If the consequences are high, scoring errors can have a severe impact, as has happened a few times in recent years. Another factor that can affect test quality has to do with the increased numbers and types of tests now being demanded. This intensified demand can lead to a shortening of the time that test producers have to develop a test. Stephen Dunbar, a co-publisher of the Iowa Test of Basic Skills, warns that this acceleration can compromise test quality.

Test Purpose and Use

Another important consideration with major impact on accuracy relates to whether or not test results are being used in appropriate ways. Assessment experts agree that tests should be designed to serve specific purposes, and should *only* be used for those purposes. If a test is going to be used for another purpose, they emphasize that it first must be demonstrated that there is evidence that the test is also valid for that new particular purpose.

For example, achievement tests are generally designed for the purpose of measuring a student’s knowledge or skills at a single point in time. However, these tests are sometimes used for other purposes. This can pose problems, especially when these include “high-stakes” purposes, such as whether or not to hold students back (retention), or whether to allow or deny entrance into special programs or schools. One large urban school district used the Iowa Test of Basic Skills (ITBS) to decide whether or not to hold students back. Subsequent studies pointed to several problems. For one thing, the test did not relate to either state or district standards—and these standards represented the content that teachers were expected to teach and students to learn. There were also issues relating to different scores on different forms of the test, so it was possible that some students who were held back would not have been if they had taken one of the other forms and vice versa. This problem would be less important if the test had



been used only for its intended purpose: to measure student achievement at a single point in time. When used to determine whether or not students should advance a grade, its impact became much more serious.

Standards-Based Instruction and Assessment

In a larger social context, the development and refinement of state standards that detail what students should know and be able to do at each grade level has led to a major shift in test design. It seems essential in today's climate of standards-based instruction that statewide tests (and other forms of assessment) should be aligned with state standards.

Test researcher Robert Linn, past president of the American Educational Research Association, advises: "Develop standards, then assessments." The reason—if the test measures different content than the state standards, then the test cannot accurately show if students are achieving the state standards. Many states are making progress toward the goal of aligning their testing programs with state standards. However, this will require continuing attention, particularly in light of the *No Child Left Behind Act*, which calls for more testing than many states now conduct. When state tests are well aligned with state standards, and when research shows that the test provides reliable information on student understanding, then such tests will present a much more accurate picture of how well students are achieving the standards.

What You Can Do

- Approach test scores with the general awareness that when tests are used for their designed purposes, they can be very helpful and reliable. At the same time, remember that they are not perfect, and that parents and citizens have every right to ask questions and investigate further.
- When reviewing a test score or school ranking based on test scores, ask the Testing Director or other administrator in charge of testing if the tests were designed for the intended purpose, especially if there are major consequences for students or schools.
- Refer to results of a number of different tests and grades whenever possible. Looking at performance from the viewpoint of these "**multiple measures**" will increase the probability that more accurate and appropriate decisions are made about students and schools.
- Read other articles about testing. Some are suggested below.

Useful Resources

How Accurate are the STAR National Percentile Rank Scores for Individual Students? An Interpretive Guide. www.cse.ucla.edu/CRESST/Reports/drrguide.pdf

AERA Position Statement Concerning High-Stakes Testing in PreK-12 Education
www.aera.net/about/policy/stakes.htm

Standards for Educational Accountability Systems
<http://www.cse.ucla.edu/cresst2/products/newsletters/polbrf54.pdf>

High Stakes: Testing for Tracking, Promotion and Graduation
<http://bob.nap.edu/html/highstakes/#summary>

Beyond Test Scores: Taking the Big-Picture View of Student Success
<http://www.asbj.com/evs/97/beyondtestscores.html>

Ron Dietel, the original author of this article, is a member of the Public Understanding strand of CAESL, and the Assistant Director for Research Use and Communications at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). CAESL Reviewers included: Jacquely Barber, Lincoln Bergman, Grace Coates, Kathy DiRanna, Joan Herman, Julia Koppich, Karen Milligan, Mike Timms, and a group of parents and teachers who provided their comments before we finalized this series of briefs.

Note: This article was developed by the Public Understanding strand of CAESL to summarize basic information for parents and the general public. It is not a CAESL position statement nor does it necessarily represent the precise views of diverse reviewers. We welcome comments!

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This material is based upon work supported by the National Science Foundation under Grant No. 0119790.

© CAESL 2003. All rights reserved. Permission to reproduce, with CAESL copyright notice included, is hereby granted.