Many people have very strong views about the role of standardized tests in education. Some believe they provide an unbiased way to determine an individual's cognitive skills as well as the quality of a school or district. Others believe that scores from standardized tests are capricious, do not represent what students know, and are misleading when used for accountability purposes. Many educational psychologists and testing experts have nuanced views and make distinctions between the information standardized tests can provide about students' performances and how the tests results are interpreted and used. In this nuanced view, many of the problems associated with standardized tests arise from their high stakes use such as using the performance on one test to determine selection into a program, graduation, or licensure, or judging a school as high vs low performing.

Are standardized tests biased?

In a multicultural society one crucial question is: Are standardized tests biased against certain social class, racial, or ethnic groups? This question is much more complicated than it seems because bias has a variety of meanings. An everyday meaning of bias often involves the fairness of using standardized test results to predict potential performance of disadvantaged students who have previously had few educational resources. For example, should Dwayne, a high school student who worked hard but had limited educational opportunities because of the poor schools in his neighborhood and few educational resources in his home, be denied graduation from high school because of his score on one test. It was not his fault that he did not have the educational resources and if given a chance with a change his environment (e.g. by going to college) his performance may blossom. In this view, test scores reflect societal inequalities and can punish students who are less privileged, and are often erroneously interpreted as a reflection of a fixed inherited capacity. Researchers typically consider bias in more technical ways and three issues will be discussed: item content and format; accuracy of predictions, and stereotype threat.

*Item content and format*. Test items may be harder for some groups than others. An example of social class bias in a
multiple choice item asked students the meaning of the term \textit{field}. The students were asked to read the initial sentence in italics and then select the response that had the same meaning of field (Popham 2004, p. 24):

\textit{My dad's field is computer graphics.}

\begin{itemize}
  \item a. \textit{The pitcher could field his position}
  \item b. \textit{We prepared the field by plowing it}
  \item c. \textit{The doctor examined my field of vision}
  \item d. \textit{What field will you enter after college?}
\end{itemize}

Children of professionals are more likely to understand this meaning of field as doctors, journalists and lawyers have "fields", whereas cashiers and maintenance workers have jobs so their children are less likely to know this meaning of field. (The correct answer is D).

Testing companies try to minimize these kinds of content problems by having test developers from a variety of backgrounds review items and by examining statistically if certain groups find some items easier or harder. However, problems do exist and a recent analyses of the verbal SAT tests indicated that whites tend to scores better on easy items whereas African Americans, Hispanic Americans and Asian Americans score better on hard items (Freedle, 2002). While these differences are not large, they can influence test scores. Researchers think that the easy items involving words that are used in every day conversation may have subtly different meanings in different subcultures whereas the hard words (e.g. vehemence, sycophant) are not used in every conversation and so do not have these variations in meaning. Test format can also influence test performance. Females typically score better at essay questions and when the SAT recently added an essay component, the females overall SAT verbal scores improved relative to males (Hoover, 2006).

\textbf{Accuracy of predictions}

Standardized tests are used among other criteria to determine who will be admitted to selective colleges. This practice is justified by predictive validity evidence— i.e. that scores on the ACT or SAT are used to predict first year college grades. Recent studies have demonstrated that the predictions for black and Latino students are less accurate than for white students and that predictors for female students are less accurate than male students (Young, 2004). However, perhaps surprisingly the test scores tend to slightly over predict success in college for black and Latino students, i.e. these students are likely to attain lower freshman grade point averages than predicted by their test scores. In contrast, test scores tend to slightly under predict success in college for female students, i.e. these students are likely to attain higher freshman grade point averages than predicted by their test scores. Researchers are not sure why there are differences in how accurately the SAT and ACT test predict freshman grades.

\textbf{Stereotype threat}

Groups that are negatively stereotyped in some area, such as women's performance in mathematics, are in danger of stereotype threat, i.e. concerns that others will view them through the negative or stereotyped lens (Aronson & Steele, 2005). Studies have shown that test performance of stereotyped groups (e.g. African Americans, Latinos, women) declines when it is emphasized to those taking the test that (a) the test is high stakes, measures intelligence or math
and (b) they are reminded of their ethnicity, race or gender (e.g. by asking them before the test to complete a brief demographic questionnaire). Even if individuals believe they are competent, stereotype threat can reduce working memory capacity because individuals are trying to suppress the negative stereotypes. Stereotype threat seems particularly strong for those individuals who desire to perform well. Standardized test scores of individuals from stereotyped groups may significantly underestimate actual their competence in low-stakes testing situations.

Do teachers teach to the tests?

There is evidence that schools and teachers adjust the curriculum so it reflects what is on the tests and also prepares students for the format and types of items on the test. Several surveys of elementary school teachers indicated that more time was spent on mathematics and reading and less on social studies and sciences in 2004 than 1990 (Jerald, 2006). Principals in high minority enrollment schools in four states reported in 2003 they had reduced time spent on the arts. Recent research in cognitive science suggests that reading comprehension in a subject (e.g. science or social studies) requires that students understand a lot of vocabulary and background knowledge in that subject (Recht & Leslie, 1988). This means that even if students gain good reading skills they will find learning science and social studies difficult if little time has been spent on these subjects.

Taking a test with an unfamiliar format can be difficult so teachers help students prepare for specific test formats and items (e.g. double negatives in multiple choice items; constructed response). Earlier in this chapter a middle school teacher, Erin, and Principal Dr Mucci described the test preparation emphasis in their schools. There is growing concern that the amount of test preparation that is now occurring in schools is excessive and students are not being educated but trained to do tests (Popham, 2004).

Do students and educators cheat?

It is difficult to obtain good data on how widespread cheating is but we know that students taking tests cheat and others, including test administrators, help them cheat (Cizek, 2003; Popham 2006). Steps to prevent cheating by students include protecting the security of tests, making sure students understand the administration procedures, preventing students from bringing in their notes or unapproved electronic devices as well as looking at each others answers. Some teachers and principals have been caught using unethical test preparation practices such as giving actual test items to students just before the tests, giving students more time than is allowed, answering students’ questions about the test items, and actually changing students’ answers (Popham, 2006). Concerns in Texas about cheating led to the creation of an independent task force in August 2006 with 15 staff members from the Texas Education Agency assigned to investigate test improprieties. (Jacobson, 2006). While the pressure on schools and teachers to have their student perform well is large these practices are clearly unethical and have lead to school personnel being fired from their jobs (Cizek, 2003).

Summary

Standardized tests are developed by a team of experts and are administered in standard ways. They are used for a variety of educational purposes including accountability. Most elementary and middle school teachers are likely to be responsible for helping their students attain state content standards and achieve proficiency on criterion-referenced
achievement tests. In order for teachers to interpret test scores and communicate that information to students and parents they have to understand basic information about measures of central tendency and variability, the normal distribution, and several kinds of test scores. Current evidence suggests that standardized tests can be biased against certain groups and that many teachers tailor their curriculum and classroom tests to match the standardized tests. In addition, some educators have been caught cheating.

**Key terms**

- Achievement tests
- Aptitude tests
- AYP (Annual Yearly Progress)
- Criterion referenced tests
- Diagnostic tests
- Frequency distribution
- Grade equivalent scores
- High stakes tests
- Mean
- Median
- Mode
- Norm referenced tests
- Range
- Standard deviation
- Stanine
- Z-score

**On the Internet**

- [http://www.cse.ucla.edu/] > The National Center for Research on Evaluation, Standards, and Student Testing (CRESST) at UCLA focuses on research and development that improves assessment and accountability systems. It has resources for researchers, K-12 teachers, and policy makers on the implications of NCLB as well as classroom assessment.

- [www.ets.org] > This is the home page of Educational Testing services which administers the PRAXIS II series of tests and has links to the testing requirements for teachers seeking licensure in each state District of Columbia and the US Virgin Islands.

- [http://www.ed.gov/nclb/landing.jhtml] > This is US Department of Education website devoted to promoting information and supporting and NCLB. Links for teachers and the summaries of the impact of NCLB in each state are provided.

**References**


