

A photograph of a classroom scene. In the foreground, a young boy with dark skin and short hair is sitting at a desk, looking towards the right. He has a pencil in his mouth. In the background, a girl with long brown hair is standing and looking towards the camera. To the right, another child's hand is raised high. The background is a chalkboard with some faint writing and numbers. The overall scene is slightly blurred, giving it a candid feel.

Chapter 15

***Understanding  
Standardized  
Assessment***



# Overview

- Standardized Tests
- Using Standardized Tests for Accountability  
Purposes: High-stakes Testing
- Standardized Testing and Technology

# Standardized Tests

- Nature of Standardized Tests
  - Designed by people with specialized knowledge and training in test construction
  - Every person who takes the test responds to the same items under the same conditions
  - The answers are evaluated according to the same scoring standards
  - The scores are interpreted through comparison to the scores obtained from a group that took the same test under the same conditions or through comparison to a predetermined standard
  - Purpose is to get an accurate and representative sample of some characteristic a person possesses

# Standardized Tests

- Uses of Standardized Tests
  - Identify students' general academic strengths and weaknesses
  - Help create instructional plans
  - Select students for inclusion in special programs

# Standardized Tests

- Criteria for Evaluating Standardized Tests
  - Reliability
    - How consistently test takers respond to test items
  - Validity
    - Accuracy of the inferences we draw about the test taker
    - Based on evidence for content validity, predictive validity, construct validity
  - Normed Excellence
    - How accurately the norm group represents the population of students for whom the test is intended
  - Examinee Appropriateness
    - Appropriateness for a particular group of students

# Reliability

- **Split-Half Reliability**
  - Administer a single test to a group of students, create two scores by dividing the test in half, and measure the extent to which the rankings change from one half to the other
- **Test-Retest Reliability**
  - Administer the same test to the same people on two occasions and measure the extent to which the rankings change over time
- **Alternate-Form Reliability**
  - Administer two equivalent forms of a test to the same group of students at the same time and compare the results

# Validity

- **Content Validity Evidence**
  - How well a test's items reflect a particular body of knowledge and skill
- **Predictive Validity Evidence**
  - How well a test predicts a student's future behavior
- **Construct Validity Evidence**
  - How well a test measures some internal attribute of a person

# Normed Excellence

- Norm Group
  - A representative sample of individuals who reflect the population of students for whom the test is intended
    - Students' scores gain meaning by being compared to scores of norm group
    - Score interpretation difficulties arise when certain groups are not represented in the norm group



# Examinee Appropriateness

- Test-Examinee Match
  - Only give tests that are appropriate for the type of decision one wants to make
    - Testing young children for admission or retention purposes is problematic because of low reliability (inconsistency of scores when retested)

# Standardized Tests

- Types of Standardized Tests
  - Achievement Tests
    - Reveal how much of a subject or skill has been learned
  - Aptitude Tests
    - Reveal how much knowledge and skill a student could acquire with effective instruction

# Types of Achievement Tests

<b>Type of Test</b>	<b>Description</b>
Single-subject achievement test	Assesses how much one has learned in a particular school subject
Achievement battery	Assesses how much one has learned in multiple school subjects
Diagnostic test	Single-subject test designed to identify specific strengths and weaknesses
Competency test	Assesses one's competence in selected basic skill areas; often taken to graduate

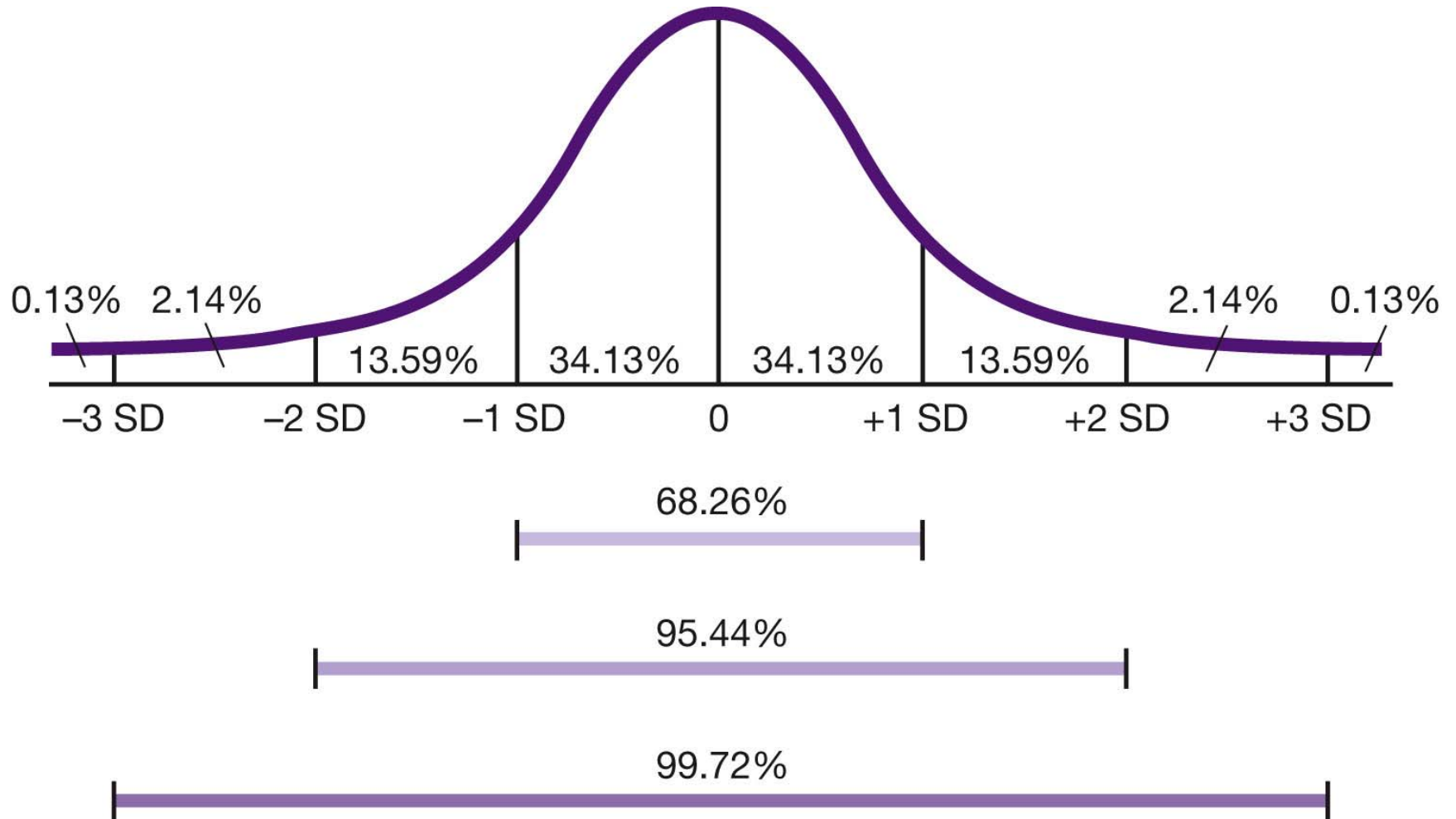
# Standardized Tests

- Types of Standardized Tests (cont'd)
  - Norm-referenced test
    - Test where one's performance is evaluated with reference to a norm group
  - Criterion-referenced test
    - Test where one's performance is evaluated with reference to some standard or criterion

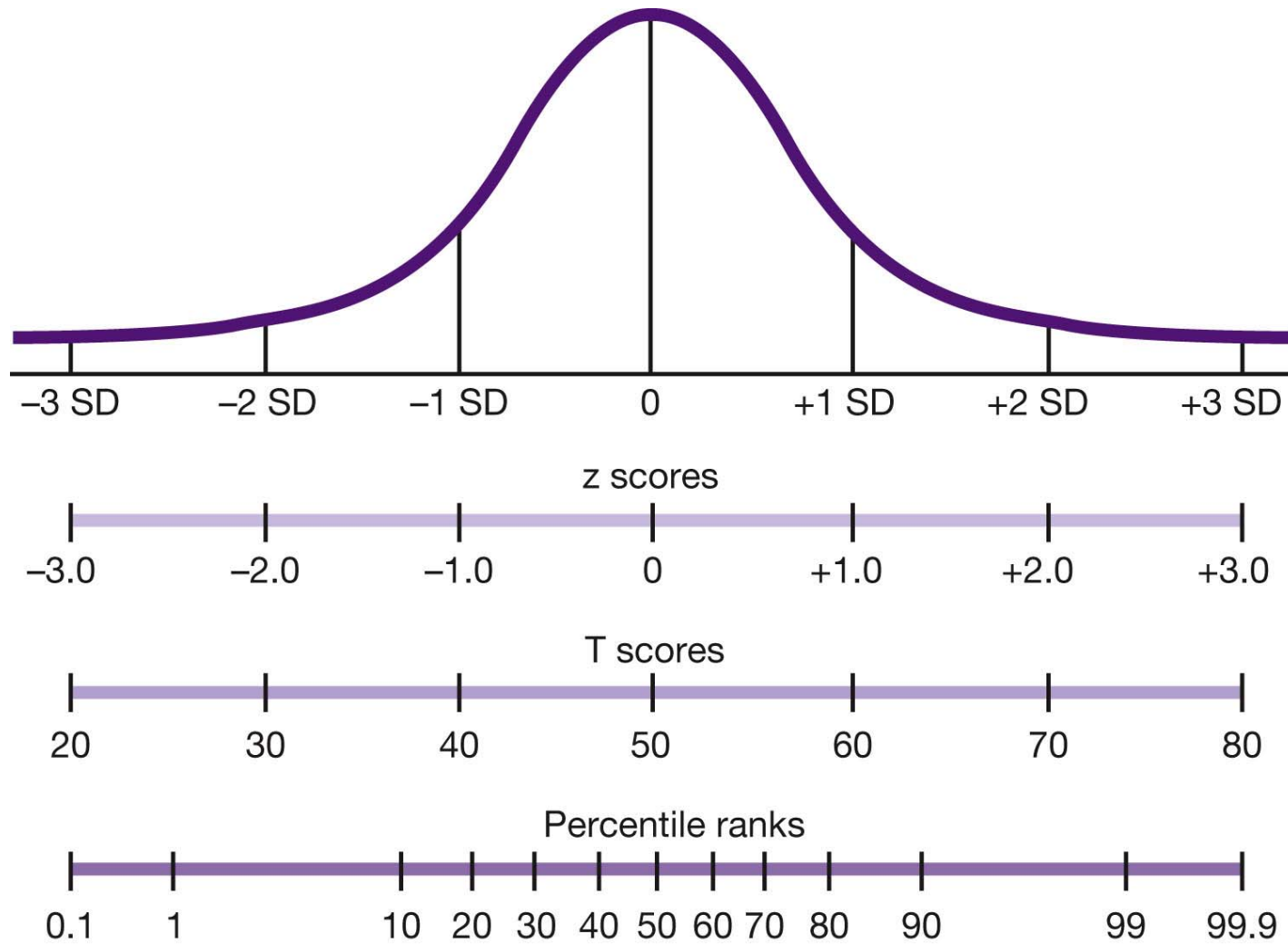
# Standardized Tests

- Interpreting Standardized Test Scores
  - Grade equivalent scores
    - Interprets test performance in terms of grade level
  - Percentile ranks
    - Score that indicates the percentage of students who are at or below a given student's score
  - Standard scores
    - Score that is expressed in terms of standard deviations, such as z score or T score
  - Stanine score
    - Type of standard score that divides a population into nine groups.

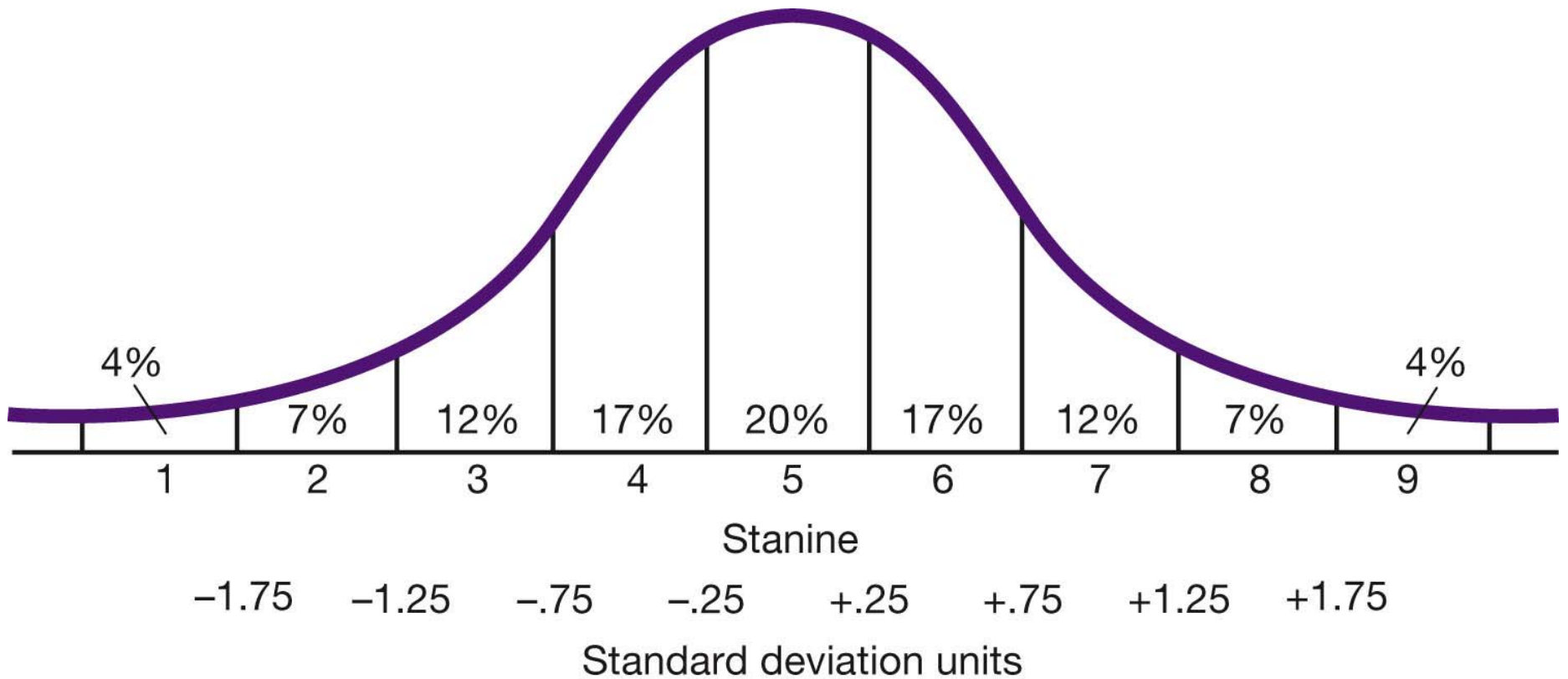
# The Normal Probability Curve



# Relationship Among z Scores, T Scores, and Percentile Ranks



# Percentage of Cases in Each Stanine





# Standardized Tests

- Misconceptions About the Nature and Use of Standardized Tests
  - A test measures what its name implies
  - All tests with the same title are the same
  - A test score accurately reflects what people know and can do
  - Two tests that claim to measure the same thing can be used interchangeably

# Standardized Tests

- Misconceptions About the Nature and Use of Standardized Tests (con't)
  - Tests are scored by adding up the number of items a person answered correctly
  - Percent correct scores are equivalent to letter grades (as in 70% = C, 80% = B, 90% = A)
  - Multiple-choice questions are useful only for measuring the lower levels of Bloom's Taxonomy
  - One can tell if a test item is useful just by making a subjective evaluation

# Using Standardized Tests for Accountability Purposes: High-Stakes Testing

- High stakes testing is...
  - Using standardized test scores, either by themselves or in conjunction with other data, to determine whether students get promoted to the next grade or graduate from high school, whether teachers and administrators receive financial rewards or demotions, and whether school districts receive additional state funds or lose their accreditation.

# Using Standardized Tests for Accountability Purposes: High-Stakes Testing

- The Push for Accountability in Education
  - 1983: National Commission on Excellence in Education publishes *A Nation at Risk: The Imperative for Educational Reform*
  - Subsequent reports critical of performance of American students on standardized tests
  - 2001: *No Child Left Behind Act*

# Video: Teacher Accountability: A Student Teacher's Perspective



# Using Standardized Tests for Accountability Purposes: High-Stakes Testing

- No Child Left Behind (NCLB)
  - Requirements of NCLB
    - Standards: States must have challenging content and achievement standards in math, reading/language arts, science
    - Testing: Annual testing of students in grades 3-12 in math, reading/language arts, science

# Using Standardized Tests for Accountability Purposes: High-Stakes Testing

- No Child Left Behind (NCLB)
  - Requirements of NCLB (cont'd)
    - Adequate yearly progress (AYP): States must demonstrate each year that an additional percentage of all students have reached the proficient level
    - Reporting: States and school districts must report how every group of students has performed on the annual assessment
    - Accountability: Rewards or punishments given to successful or failing schools

# Using Standardized Tests for Accountability Purposes: High-Stakes Testing

- No Child Left Behind (NCLB)
  - Problems Implementing NCLB
    - States with “backloaded” schedules unlikely to meet goal of 100% proficiency by 2014
    - Difference between low- and high-scoring states likely due to different standards for proficiency
    - Some states allowed larger minimum subgroup size
    - Law ignores long-standing learning problems of minorities, English language learners, and students with disabilities
    - Students with disabilities required to meet same standards as students without disabilities despite requirements of IDEA



# Using Standardized Tests for Accountability Purposes: High-Stakes Testing

- No Child Left Behind (NCLB)
  - Modifications of NCLB
    - Schools in which all students meet requirements except for one subgroup need not be labeled “needs improvement”

# Using Standardized Tests for Accountability Purposes: High-Stakes Testing

- Arguments in Support of High-Stakes Testing Programs
  - Goal Clarity
    - Students, teachers, administrators, parents better understand content and performance standards
  - Improved Quality Control
    - Progress of underserved groups is monitored
  - Beneficial Effects for Teaching
    - Increased use of effective teaching methods, focus on material related to standards, collaboration among teachers
  - Beneficial Effects for Students
    - Increased motivation to work harder to avoid penalties

# Using Standardized Tests for Accountability Purposes: High-Stakes Testing

- Arguments Critical of High-Stakes Testing Programs
  - Structural Limitations
    - Contradiction between goals of challenging standards and 100% proficiency
    - Most state assessments provide little or no useful diagnostic information to teachers
    - Test content does not reflect all views on valuable educational standards
    - States have more content standards than teachers can cover or students can master
    - Focus on outputs at the expense of inputs and processes
    - Schools that recorded gains too small to satisfy AYP are still labeled as “needs improvement”

# Using Standardized Tests for Accountability Purposes: High-Stakes Testing

- Arguments Critical of High-Stakes Testing Programs (cont'd)
  - Misinterpretation and Misuse of Test Results
    - Test results cannot be used to draw inferences about the quality of a school or a teacher's effectiveness
  - A One-Size-Fits-All Approach to Motivation
    - Rewards and punishments are insufficient bases for motivating teachers and students
  - A One-Size-Fits-All Approach to Standards
    - Students of different abilities and backgrounds should be held to different standards

# Using Standardized Tests for Accountability Purposes: High-Stakes Testing

- Arguments Critical of High-Stakes Testing Programs (cont'd)
  - Undesirable Side Effects
    - Curriculum will emphasize basic math and reading/language arts at the expense of other subjects
    - Direct instruction will supplant constructivist and inquiry-based instruction
    - States will lower their standards as much as possible

# Using Standardized Tests for Accountability Purposes: High-Stakes Testing

- Research on the Effects of High-Stakes Testing
  - Effect on Achievement
    - Overall, positive but small effects on math & reading
    - Little change in the gap between White and minority students
  - Effect on Motivation
    - In some cases, students are working harder
  - Effect on Teachers and Teaching
    - Increases in teacher-directed instruction, teaching to the test, and test prep, but some use of student-centered instruction as well
    - Increased time and resources given to “bubble kids”

# Using Standardized Tests for Accountability Purposes: High-Stakes Testing

- Research on the Effects of High-Stakes Testing (cont'd)
  - Effect on the Curriculum
    - Most schools increased time devoted to math and reading at the expense of social studies, art, music, physical education, and recess
  - Effect on the Dropout Rate
    - Hard to tell because of no common definition, but high-stakes tests appear to increase the dropout rate
  - Effect on State Standards and Test Quality
    - Some states lowered standards and made tests easier, other states did not
  - Use of Supplemental Educational Services
    - Only about 17% of eligible students participate

# Using Standardized Tests for Accountability Purposes: High-Stakes Testing

- Recommendation for Improving High-Stakes Testing
  - Each state should adopt only those content standards that represent the most important knowledge and skills that students need to learn
  - A state's content standards should clearly describe exactly what is being assessed so teachers can create lessons that directly address those standards
  - Scores on a state assessment should be reported on a standard-by-standard basis for each student, school, and district
  - States should provide school districts with additional assessment procedures to assess those standards that the required assessment does not cover
  - States should monitor the curricula of school districts to ensure that instruction addresses all content standards and subjects, not just those assessed by the required state test



# Using Standardized Tests for Accountability Purposes: High-Stakes Testing

- Recommendations for Improving High-Stakes Testing (cont'd)
  - State assessments should be designed so that all students have an equal opportunity to demonstrate which of the state's standards they have mastered
  - All tests should satisfy the standards of professional organizations
  - Teachers and principals should receive professional development training that helps them use test results to optimize children's learning
  - States should be required to have their standards and assessment instruments reviewed
  - States should be able to use improving test scores as a way to satisfy AYP

# Standardized Testing and Technology

- Using Technology to Prepare Students for Assessments
  - Some state web sites provide content and performance standards, examples of items, practice tests, and test preparation programs of for-profit companies
- Using Technology to Assess Mastery of Standards
  - Use of Supplemental Educational Services Computer-based testing programs
- Using Technology to Promote Mastery of Standards
  - To prepare for state assessments, students take computer-based instruction courses that are aligned with state standards
- Computer Adaptive Testing
  - Level of difficulty of test items geared to student's ability level