SCHOOLING
VOLUME 1, NUMBER 1, 2010

Measurement and Assessment in Schools

Fred C. Lunenburg
Sam Houston State University

ABSTRACT

Every comprehensive school assessment program can and should provide information that enhances instruction and promotes learning. In this article, I discuss the purposes of assessment, information criteria needed to make decisions about testing, creating effective tests, and the major components of a school-wide testing program.

Testing, evaluating, and measuring student progress is a part of every comprehensive assessment program in schools. Practically every member of a school faculty is involved in assessment. Teachers spend a great deal of time testing, measuring, and evaluating their students, as do counselors, social workers, and school psychologists. Few people who work in schools would deny that the modern school could operate effectively without some means of measuring and evaluating student progress, particularly in light of the *No Child Left Behind Act of 2001* (NCLB), which requires schools to meet Adequate Yearly Progress (AYP) standards in reading and math. Those schools failing to meet the standards for several years could undergo major restructuring efforts.

Adequate Yearly Progress standards require schools and districts to have (a) the same high standards of academic achievement for all; (b) statistically valid and reliable tests; (c) continuous and substantial academic improvement for all students; (d) separate, measurable annual objectives for achievement for all students, racial/ethnic groups, economically disadvantaged students, students with disabilities (IDEA, § 602), and students with limited English proficiency; and (e) graduation rates for high school and one other indicator for other schools. NCLB 1111 (b) (2) (C) (vi) defined the percentage of students who graduate from secondary school with a regular diploma in the standard number of years, and the regulation clarifies that alternate definitions that accurately measure the graduation rate are permissible. For AYP, each group of students must meet or exceed the established statewide annual objective exception: The number below proficient is reduced 10% from the prior rate, and the subgroup must make progress on other indicators; and for each group, 95% of students enrolled participate in the assessments on which AYP is based (see your state's guidelines for AYP). In this paper, I examine the purposes of assessment, establishing information criteria, creating effective tests, and the components of a testing program

Purposes of Assessment

The basic purpose of assessment is to help the student in school. More specifically, six basic purposes of assessment include: (a) to help the student understand herself; (b) to provide information for educational and vocational counseling; (c) to help administrators, faculty, and other personnel to understand the nature of their student population; (d) to evaluate the academic progress and personal development of students; (e) to help the administrative staff appraise the educational program, and (f) to facilitate curriculum revision (Fiore, 2012; Wright, 2011). Others have suggested three basic purposes of testing students: (a) to make instructional management decisions, (b) to make decisions about screening students, and (c) to make program decisions (Aiken, 2011; Guion, 2012). (See Table 1.)

Table 1 describes the relationships among general purposes: assessment context, types of decisions, type of data needed, and appropriate decision makers. For example, diagnosis of student strengths and weaknesses is one of the decision areas listed under "Instructional Management Decisions." Because the purpose of testing in this case is to make decisions about individual students, any test that is adopted must provide information about specific skills and/or subject matter understanding.

Table 1
Summary of Various Purposes for Testing

Assessment Context	Types of Decisions	Type of Data Needed	Decision Makers					
Instructional Management Decisions			**Students**	**Parents**	**Teachers**	**Adminis-trators**	**Coun-selors**	**Public**
Diagnosis	Decide students' strengths and weaknesses on specific skills	Individual student data on level of development of specific skills	✓		✓			
Placement	Place student into next most appropriate level of instruction	Scores that place students on relevant knowledge or skill continuum			✓	✓	✓	
Guidance and Counseling	Decide probability of success and satisfaction in given program of educational or vocational development	Data reflecting level of educational development of individual student relative to other students	✓	✓				✓
Student Screening Decisions								
Selection	Decide which students to be selected into or out of a program	Data that rank order individual students on relevant knowledge or skill scale			✓	✓	✓	
Certification	Determine mastery or nonmastery of specified body of knowledge or set of skills	Data reflecting individual student mastery of specified body of knowledge or set of skills			✓	✓		✓

Table 1 (Continued)

Assessment Context	Types of Decisions	Type of Data Needed	Decision Makers					
Program Decisions			**Students**	**Parents**	**Teachers**	**Adminis-trators**	**Coun-selors**	**Public**
Survey	Make educational policy decisions: determine educational development of student group	Group achievement data gathered cyclically to show trends				✓		✓
Formative Evaluation	Decide program components in need of modification	Interim and final program outcomes attained and not attained by participating subtends considered as a group			✓	✓		
Summative Evaluation	Determine if program is to be adopted, expanded, or discontinued	Program outcomes attained and not attained by participating students considered as a group			✓	✓		✓

Establishing Information Criteria

After determining the purposes of testing students, the next step is to identify the type of information needed to make the desired decisions. The information provided by any testing program can be classified into one or more of three categories or domains: (a) the *affective domain***,** which refers to attitudes, feelings, interests, and values (Krathwohl, Bloom, & Masia, 1964); (b) the *psychomotor domain***,** which refers to those skills involving neuromuscular coordination such as handwriting skills and athletic skills (Harrow, 1972); and (c) *cognitive domain***,** which is identified in a variety of ways and is usually classified along a continuum. Knowledge (that is, knowledge of facts, rules, and sequences) is viewed as a lower order cognitive skill, and higher order skills include the ability to classify, to recognize relationships, to analyze, to synthesize, and to evaluate (Bloom, 1956).

For example, if the purpose of testing is to make decisions about placement of students into academic classes of differing ability levels, it will be necessary to collect data that will yield information about how students perform on specific cognitive objectives. However, if a program has explicit affective goals, and the decision is to determine how well those goals have been reached, it will be necessary to develop or purchase measures of student attitudes, feelings, values, or interests. Both teacher-constructed and standardized instruments can be used for this purpose.

Should tests be authentic simulations of how knowledge is tested in adult work and civic settings? Many educators believe so (Guion, 2012; Odendahl, 2011; Wiggins, 2008). Performance assessment, then, calls on test makers to be creative designers not just technicians.

Creating Effective Tests

Grant Wiggins (2008) offers the following eight basic design criteria as assistance to test designers.

- Assessment tasks should be, whenever possible, authentic and meaningful-worth mastering.
- The set of tasks should be a valid sample from which apt generalizations about overall performance of complex capacities can be made.
- The scoring criteria should be authentic, with points awarded or taken off for essential successes and errors, not for what is easy to count or observe.
- The performance standards that anchor the scoring should be genuine benchmarks, not arbitrary cut scores or provincial school norms.
- The context of the problems should be rich, realistic, and enticing-with the inevitable constraints on access to time, resources, and advance knowledge of the tasks and standards appropriately minimized.
- The tasks should be validated.
- The scoring should be feasible and reliable.
- Assessment results should be reported and used so that *all* customers for the data are satisfied.

Components of a Testing Program

A comprehensive school-wide testing program begins in kindergarten and ends in the 12th grade. Tests administered throughout the school years include: assessment of emerging reading, general learning readiness, tests of general intelligence, achievement, and aptitude and interest. A typical testing program in a school district might resemble the one depicted in Table 2 (Aiken, 2011; American Society for Testing and Materials, 2012; Hattie, 2012; Murphy, 2011).

Table 2
Comprehensive School-wide Testing Program

Grade	Type of Test
K	Reading readiness
1 or 2	Learning readiness
	Reading ability
	Mental ability
3, 4, or 5	Achievement battery (language skills, including reading, mathematics, social studies, science)
6, 7, or 8	Mental ability (repeated at entrance to middle school or junior high)
	Multifactor aptitude
9, 10, or 11	Achievement battery
11 or 12	College aptitude
	Interests—personal
	Interests—vocational

In discussing the evaluation of a student's growth in schools, Popham (2010a, b) referred to three areas of measurement: (a) knowledge and understanding, (b) skills and competence, and (c) aptitude and interest. For each of these, he refers to a number of educational objectives and the appropriate means of evaluation. It would seem, therefore, that every school should have at least the following components in the way of a testing battery.

Emerging Reading Test

An emerging reading test should be administered in kindergarten or first grade to determine the child's readiness to profit from reading. Examples include the Gates-MacGinitie Reading Tests (Houghton-Mifflin), Lee-Clark Emerging Reading Test (California Test Bureau), and the Murphy-Durrell Emerging Reading Analysis (Harcourt Brace Jovanovich). These tests measure speed and accuracy, vocabulary, comprehension, and similarities and differences in printed letters.

Learning Readiness Test

A learning readiness test should be administered in grade one or two demonstrating the ability to mark pictures and letters, and to identify words that match given ones. Examples include the Metropolitan Readiness Test (a group test that assesses six important aspects of readiness for formal first-grade instruction: word meaning, listening, visual perception, alphabet, numbers, and copying); and Primary Mental Abilities Test (measures verbal meaning, number facility, reasoning, perceptual speed, and spatial relations).

Intelligence Test

An intelligence test should be administered in grade one, again toward the end of the elementary grades, and again early in high school. An individual measure is better than a group measure. There is a wide discrepancy between tests concerning the accurate measurement of a child's intelligence. Among the best known intelligence tests are the California Short Form Test of Mental Maturity, Lorge-Thorndike Intelligence Test, Otis-Lennon Mental Ability Test, Revised Stanford-Binet Scale, and the Wechsler Intelligence Scales.

Achievement Tests

At a very minimum, an achievement test battery including language, reading, mathematics, and social studies should be given periodically during the student's twelve years of public school education. Examples include the Comprehensive Tests of Basic Skills, Iowa Tests of Basic Skills, Sequential Tests of Educational Progress, and Stanford Achievement Tests. Generally, these tests measure the capacity to comprehend written material, to think scientifically and analytically, and to display some understanding of the process of history. A norm-referenced achievement test, such as those mentioned above, should be administered in elementary school and high school in addition to the state's mandated achievement test.

Interest and Aptitude Tests

Measures of interests and aptitudes might be administered periodically to individuals, or groups of students, for purposes of placement or selection. Examples of aptitude measures include the Differential Aptitude Tests, School and College Ability Test, American College Testing Program, and the Scholastic Aptitude Test. Generally, these tests measure basic verbal and mathematical ability and reasoning.

Perhaps the best reference available for a quick summary and review of nearly all tests on the market is the *Mental Measurements Yearbook* (Spies, 2008). It should be noted, however, that not all of the tests that are published are available for use without the proper training. Many universities provide courses that are concerned with the understanding, interpretation, and use of one specific test.

There may be missing elements in the conceptualization on which standard test theory is based. Those elements are models for just how people know what they know and do what they can do, and the ways in which they increase these capacities. Different models are useful for different purposes; therefore, broader or alternative student models are being proposed by test experts. For example, test experts from Educational Testing Service consider a variety of directions in which standard test theory might be extended. They discuss the role of test theory in light of recent work in cognitive and educational psychology, test design, student modeling, test analysis, and the integration of assessment and instruction.

Conclusion

Every comprehensive school assessment program involves testing, evaluating, and measuring student progress. In this article, I discussed a school-wide assessment program, including the purposes of assessment, information criteria needed to make decisions about testing, creating effective tests, and the major components of a testing program. The components of a school-wide testing program should include the following tests: emerging reading test, learning readiness test, intelligence test, achievement tests, and interest and aptitude tests.

References

Aiken, L. R. (2011). *Psychological testing and assessment.* Boston, MA: Allyn & Bacon.

American Society for Testing and Materials. (2012). *Nondestructive testing*. West Conshohocken, PA: Author.

Bloom, B. S. 1956. *Taxonomy of educational objectives, handbook I: Cognitive domain*. New York, NY: Longman.

Fiore, L. B. (2012). *Assessment of young children*. New York, NY: Routledge.

Guion, R. M.(2012). *Assessment, measurement, and prediction for personnel decisions* (2nd ed.). New York, NY: Taylor & Francis.

Harrow, A. J. (1972). *A taxonomy of the psychomotor domain*. New York, NY: Longman.

Hattie, J. (2012). *Intelligence and intelligence testing*. New York, NY: Routledge.

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives, handbook II: Affective domain*. New York, NY: Longman.

Murphy, L. L. (2011). *Tests in print VIII*. Lincoln, NE: Boros Institute of Mental Measurements.

Odendahl, N. (2011). *Testwise: Understanding educational assessment, volume one*. Lanham, MD: Roman & Littlefield.

Popham, J. W. (2010a). *Classroom assessment: What teachers need to know*. Upper Saddle River, NJ: Prentice Hall.

Popham, J. W. (2010b). *Educational assessment: What school leaders need to know and understand*. Thousand Oaks, CA: Sage.

Spies, R. A. (2008). *Mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Wiggins, G. P. (2008). *Schooling by design: Mission, action, and achievement. Alexandria, VA: Association for Supervision and Curriculum Development.*

Wright, J. (2011). *Conducting psychological assessments*. New York, NY: Wiley.