

SABES Math Bulletin

..... Building Research Into Practice
..... Volume 2, Issue 2 January, 2008

In this issue we continue our exploration of A Review of the Literature in Adult Numeracy: Research and Conceptual Issues, published in March, 2006. We focus on the “Assessment” section. From No Child Left Behind to the NRS (National Reporting System) standards, every educator’s attention has been directed to think about assessment. What is educational assessment? According to the System for Adult Education Support (SABES) in Massachusetts, assessment is:

Any systematic method of obtaining information from tests and other sources, used to draw inferences about characteristics of people, objects, or programs; the process of gathering, describing, or quantifying information about performance; an exercise-such as a written test, portfolio, or experiment-that seeks to measure a student’s skills or knowledge in a subject area.

AIR (American Institutes for Research) which was commissioned by the U.S. Department of Education to perform the literature review, examined adult education assessment literature and practices then presented information about the nature and purposes of assessment, a conceptualization of ‘good’ assessments and an outline of several widely used instruments. We share an abbreviated version of this report segment with you below. In addition, we share some definitions of key assessment terms and sites to seek further information if you become inspired. Finally, we include some information from the National Council of Teachers of Mathematics (NCTM) regarding their “Assessment Principle.”

Enjoy! And think about your role as a teacher or adult education administrator and how this work informs and responds to numeracy assessment issues in today’s world.

Inside This Issue

AIR on Assessment Issues in Adult Numeracy.....	1
SABES Assessment Glossary Terms	3
NCTM Assessment Perspectives.....	7
National Assessment Programs.....	8

AIR on Assessment Issues in Adult Numeracy

Research on numeracy assessment is “limited and patchy,” according to the AIR report authors (AIR, 2006 p.36). For instance, the authors note that in its last three conferences, ALM (Adults Learning Mathematics) showcased 88 presentations, in which only four discussed assessment. They explain that “numeracy is often not discussed as a stand-alone topic” in adult education assessment-related research and papers (Ibid., p.35). However, the AIR literature review noted that numeracy tasks are often included in assessment procedures that are undertaken for three main purposes: program entrance, teaching and learning, and program exit stage. But with the introduction of the National Reporting

Continued on page 2

System (NRS) accountability mandates, emphasis has been on the use of “standardized test(s) to measure educational gain in ABE and ESL students, which puts most emphasis on assessment for entrance and exit from programs or program levels. The inference here is that assessment for instructional purposes, during the learning and teaching stages, has taken a back seat.

Whereas NCTM (see article p 7 and many adult educators had been emphasizing authentic assessments (see glossary terms, p. 3), the NRS requirement that states use standardized tests that are reliable and valid has resulted in less development of alternative assessment tools and methodologies.

In fact, two tests have surfaced as the primary tools used to assess student gains and often their entry status into adult education programs: the TABE and CASAS. The report describes each of the tests, and even a quick read of the data will reveal author support for the CASAS over the TABE. However, they note that for both tests “various concerns and questions can be raised that touch on validity, construct coverage, permissible interpretations of the meaning of test scores, and alignment between assessments and instruction.”

(Ibid., p. 40) Furthermore, the authors state that both tests fail to “adequately address conceptions of numeracy in the integrative (mathematical) phase,” where students must demonstrate problem solving and critical thinking skills. (Ibid., p. 40)

“Good assessments,” as observed in the literature...need to enable learners, teachers, and programs to identify the ability to transfer and apply learned numeracy skills in real, functional contexts, but not those where the mathematics is explicit and obvious. At the same time, the ability to handle functional numeracy tasks will not necessarily imply that a learner has acquired generalizable skills or the ability to handle abstract or more formal mathematical concepts and ideas. Thus, a mix of both functional and more abstract tasks may be needed. (Ibid, p. 47)

The Reports’ authors conclude that “teachers have a key role in changing assessment practices.” Teachers, they say, know best the goals of mathematics learning for adults and how well assessment practices are aligned with such goals.” It is teachers, they assert, whose knowledge, values, decisions, practices, and training need to be considered when thinking of improving assessments in ways that can better address instruction, validity, and accountability needs.” (Ibid, p. 48)

Additional Assessment Resources

CRESST/UCLA, Assessment Glossary.
www.cse.ucla.edu/CRESST/pages/glossary.htm

CTB McGraw-Hill, Glossary of Assessment Terms.
www.ctb.com/articles/

Joint Committee on Standards for Educational and Psychological Testing. (1999). Standards for Educational and Psychological Testing. Washington, DC: AERA/APA/NCME.

SABES Assessment Glossary Terms

<http://www.sabes.org/assessment/glossary.htm>

Alternative Assessment (Authentic Assessment, Performance Assessment)

An assessment that requires students to generate a response to a question rather than choose from a set of responses provided to them. Exhibitions, investigations, demonstrations, written or oral responses, journals, and portfolios are examples of the assessment alternatives we think of when we use the term “alternative assessment.” Ideally, alternative assessment requires students to actively accomplish complex and significant tasks, while bringing to bear prior knowledge, recent learning, and relevant skills to solve realistic or authentic problems.

Analytic scoring

Evaluating student work across multiple dimensions of performance rather than from an overall impression (holistic scoring). In analytic scoring, individual scores for each dimension are scored and reported. For example, analytic scoring of a history essay might include scores of the following dimensions: use of prior knowledge, application of principles, use of original source material to support a point of view, and composition. An overall impression of quality may be included in analytic scoring.

Anchor tests

1. A common set of items administered with each of two or more different forms of a test for the purpose of equating the scores obtained on these forms.
2. “Anchor papers provide a connection between a rubric narrative and student writing, and an example of what writing at a certain score of the rubric should look like.” (From: the REEP Writing Assessment Trainer’s Manual.) (see also Benchmark tasks; Benchmarking.)

Assessment

Any systematic method of obtaining information from tests and other sources, used to draw inferences about characteristics of people, objects, or programs; the process of gathering, describing, or quantifying information about performance; an exercise—such as a written test, portfolio, or experiment—that seeks to measure a student’s skills or knowledge in a subject area.

Continued on page 4

Authentic assessment (alternative or performance)

Assessment is authentic when we directly examine student performance on worthy intellectual tasks....Authentic assessments present the student with a full array of tasks that mirror the priorities and challenges found in the best instructional activities.

Benchmark

A detailed description of a specific level of student performance expected of students at particular ages, grades, or development levels. Benchmarks are often represented by samples of student work. A set of benchmarks can be used as “checkpoints” to monitor progress toward meeting performance goals within and across grade levels. In ABE, SPLs (Student Performance Levels) are examples of benchmarks; targets for instruction.

Benchmark tasks

Pieces of student work selected by a group of lead teachers as exemplifying a certain score level. (See also Anchor Test.)

Benchmarking

Comparing performances of people on the same task; raters use “anchors” to score student work, usually comparing the student performance to the “anchor”; benchmarking is a common practice in the business world. (See also Anchor Test.)

Bias

A situation that occurs in testing when items systematically measure differently for different ethnic, gender, or age groups. Test developers reduce bias by analyzing item data separately for each group, then identifying and discarding items that appear to be biased.

Competency

A group of characteristics, native or acquired, which indicate an individual’s ability to acquire skills in a given area.

Competency-based assessment (criterion-referenced)

Measures an individual’s performance against a predetermined standard of acceptable performance. Progress is based on actual performance rather than on how well learners perform in comparison to others. CASAS and BEST are examples of competency-based assessments.

Constructed response item

An exercise for which examinees must create their own responses or products (performance assessment) rather than choose a response from an enumerated set (multiple choice).

Criteria

Guidelines, rules, characteristics, or dimensions that are used to judge the quality of student performance. Criteria indicate what we value in student responses, products or performances. They may be holistic, analytic, general, or specific. Scoring rubrics are based on criteria and define what the criteria mean and how they are used.

Criterion-referenced assessment (competency-based assessment)

An assessment where an individual’s performance is compared to a specific learning objective or performance standard and not to the performance of other students. Criterion-referenced assessment tells us how well students are performing on specific goals or standards rather than just telling how their performance compares to a norm group of students nationally or locally. In criterion-referenced assessments, it is possible that none, or all, of the examinees will reach a particular goal or standard.

Criterion-referenced tests

Tests that assess a learner’s achievement against an absolute standard or criterion of performance (rather than against a norming group).

Continued on next page

SABES Assessment Glossary

Continued from page 4

Cut score

A specified point on a score scale, such that scores at or above that point are interpreted or acted upon differently from scores below that point. (See also Performance Standard.)

Evaluation

When used for most educational settings, evaluation means to measure, compare, and judge the quality of student work, schools, or a specific educational program.

Formative assessment

Assessment that provides feedback to the teacher for the purpose of improving instruction.

Grade Level Equivalent (GLE)

The school grade level for a given population for which a given score is the median score in that population. For example, if a test was administered during the month of October to a norming group of sixth grade students and the median scale score obtained was 475, then the grade equivalent for a scale score of 475 on that test would be set at 6.1 - 6 representing Grade 6 and .1 representing the month of October (September is taken as the beginning of the school year and equals .0).

Grade level norms

Interpreting scores on a test in reference to the average performance of children at each grade level. The TABE and the ABLE provide norms for adults in ABE programs that permit test users to interpret scores both in grade levels and in relation to adult performance on the tests.

High-stakes test

A test used to provide results that have important, direct consequences for examinees, programs, or institutions involved in the testing. For example, MCAS (K-12) is consid-

ered a high-stakes test because children who do not pass the examination do not receive a high school diploma, regardless of their performance in other areas of their school education.

Holistic scoring

Evaluating student work in which the score is based on an overall impression of student performance rather than multiple dimensions of performance (analytic scoring).

Inter-rater reliability

The consistency with which two or more judges rate the work or performance of test takers.

Item response theory (IRT)

A method for scaling individual items for difficulty in such a way that an item has a known probability of being correctly completed by an adult of a given ability level.

Iterative

A term used in research to refer to the repetition of a cycle of processes with an eye toward moving ever more closely toward desired results.

Materials-based assessment

Evaluation of learners on the basis of tests following the completion of a particular set of curriculum materials. A commercial text and its accompanying workbook is an example of this type of assessment.

Measurement

Process of quantifying any human attribute pertinent to education without necessarily making judgements or interpretations.

Norm-referenced assessment

An assessment where student performance(s) are compared to those of a larger group. Usually the larger group or "norm group" is a national sample representing a wide and diverse cross-section of students. Students, schools, districts, and even states are compared or rank-ordered in relation to the norm group. The purpose

Continued on next page

SABES Assessment Glossary

Continued from page 5

is usually to sort students, not to measure achievement towards some criterion of performance.

Norm-referenced test

An objective test that is standardized on a group of individuals whose performance is evaluated in relation to the performance of others; contrasted with criterion-referenced test. Most standardized achievement tests are referred to as norm-referenced.

Norms

A performance standard that is established by a reference group and that describes average or typical performance. Usually norms are determined by testing a representative group and then calculating the group's test performance. Participatory assessment (alternative, authentic performance-based) A process for examining performance that views literacy as practices and critical reflection; requires the use of a broad range of strategies in assessment; and provides an active role for learners in the assessment process.

Performance assessment (alternative, authentic, participatory)

Performance assessment is a form of testing that requires students to perform a task rather than select an answer from a ready-made list. Performance assessment is an activity that requires students to construct a response, create a product, or perform a demonstration. Usually there are multiple ways that an examinee can approach a performance assessment and more than one correct answer.

Performance standards

I. A statement or description of a set of operational tasks exemplifying a level of performance associated with a more general content standard; the statement may be used to guide

judgments about the location of a cut score on a score scale; the term often implies a desired level of performance. 2. Explicit definitions of what students must do to demonstrate proficiency at a specific level on the content standards.

Performance task

A carefully planned activity that requires learners to address all the components of performance of a standard in a way that is meaningful and authentic. Performance tasks can be used for both instructional and assessment purposes.

Portfolio assessment

A portfolio is a collection of work, usually drawn from students' classroom work. A portfolio becomes a portfolio assessment when (1) the assessment purpose is defined; (2) criteria or methods are made clear for determining what is put into the portfolio, by whom, and when; and (3) criteria for assessing either the collection or individual pieces of work are identified and used to make judgments about performance. Portfolios can be designed to assess student progress, effort, and/or achievement, and encourage students to reflect on their learning.

Rating scales

Values given to student performance. Subjective assessments are made on predetermined criteria for documenting where learners fall on a continuum of proficiency. Rating scales include numerical scales or descriptive scales.

Reliability

How accurately a score will be reproduced if an individual is measured again. The degree to which the results of an assessment are dependable and consistently measure particular student knowledge and/or skills. Reliability is an indication of the consistency of scores across raters, over time, or across different tasks or items that measure the same thing.

Continued on next page

SABES Assessment Glossary

Continued from page 6

Rubrics

Specific sets of criteria that clearly define for both student and teacher what a range of acceptable and unacceptable performance look like. Criteria define descriptors of ability at each level of performance and assign values to each level.

Scale scores

A score to which raw scores are converted by numerical transformation (e.g., conversion of raw scores to percentile ranks or standard scores); units of a single, equal-interval scale that are applied across levels of a test; for example, on TABE 7 & 8, scale scores are expressed as numbers that may range from 0 through 999.

Selected response item

An exercise for which examinees must choose a response from an enumerated set (multiple choice) rather than create their own responses or products (performance assessment).

Standard error of measurement

The difference between an observed score and the corresponding true score or proficiency; the standard deviation of an individual's observed scores from repeated administrations of a test or parallel forms of a test, under identical conditions. Because such data cannot generally be collected, the standard error of measurement is usually estimated from group data.

Standardization

A consistent set of procedures for designing, administering, and scoring an assessment. The purpose of standardization is to assure that all students are assessed under the same conditions so that their scores have the same meaning and are not influenced by differing conditions. Standardized procedures are very important when scores will be used to compare individuals or groups.

Standardized testing

A test designed to be given under specified, standard conditions to obtain a sample of learner behavior that can be used to make inferences about the learner's ability. Standardized testing allows results to be compared statistically to a standard such as a norm or criteria. If the test is not administered according to the standard conditions, the results are invalid.

Student performance level (SPL)

A standard description of a student's (ESOL) language ability at a given level in terms of speaking, listening, reading, writing, and the ability to communicate with a native speaker; a profile of skill levels for a student can thus be assigned and used for placement, instructional, or reporting purposes.

Summative assessment

A culminating assessment, which gives information on students' mastery of content, knowledge, or skills.

Triangulation

A process of combining methodologies to strengthen the reliability of a design approach; when applied to alternative assessment, triangulation refers to the collection and comparison of data or information from three different sources or perspectives.

Validity

The extent to which an assessment measures what it is supposed to measure and the extent to which inferences and actions made on the basis of test scores are appropriate and accurate. A valid standards-based assessment is aligned with the standards intended to be measured, provides an accurate and reliable estimate of students' performance relative to the standard, and is fair. An assessment cannot be valid if it is not reliable.

NCTM Assessment Perspectives

NCTM Assessment Principle

The National Council of Teachers of Mathematics offers a succinct assessment principle:

Assessment should support the learning of important mathematics and furnish useful information to both teachers and students.

NCTM envisions quality assessment as an “integral” part of classroom practice, “a routine part of the ongoing classroom activity rather than an interruption.” (NCTM, Principles and Standards for School Mathematics, 2000, p.23) Assessment, they say, “(should) not merely be done to students; ...it should be done for students, to guide

and enhance their learning.” (Ibid, p.22) Furthermore, they clearly and strongly advocate for use of a variety of assessment methods; and they link assessment to teachers’ instructional decision-making.

NCTM firmly objects to over-reliance on testing, especially individual, time-limited, pencil and paper tasks. They warn that such over-reliance “may give an incomplete and perhaps distorted picture of students’ performance.” To avoid this distortion, they recommend teachers “look for a convergence of evidence from different sources.” See chart below for a listing of sources.

Assessment Tool or Method	Purpose/Instructional Goal
Open-ended questions	Show students’ preferred methods of approaching problems or ability to reason in different ways about a problem
Constructed-response tasks	Determine if students can apply procedures
Selected-response items	Determine if students can apply procedures
Performance tasks	Illuminate students’ ability to apply mathematics in complex or new situations
Observations	Allow insight into students’ thinking; monitor changes
Conversations	Allow insight into students’ thinking
Journals	Monitor changes in students’ thinking and reasoning over time
Portfolios	Monitor changes in students’ thinking and reasoning over time

As teachers gather information from a variety of sources, they become better prepared to make instructional decisions, according to NCTM. They know better what to review, what to re-visit, what to move on to, and what parts of instruction made sense to students and what parts did not. The assemblage of information using a variety of assessment techniques also provides a clearer picture of each individual student’s understanding of and approach to mathematics. And the point of assessment, they emphasize, is this deeper understanding of students, which

results in more effective instructional decision-making. (Ibid., p. 24)

“Teachers must understand their mathematical goals deeply, they must understand how their students may be thinking about mathematics, they must have a good grasp of possible means of assessing students’ knowledge, and they must be skilled in interpreting assessment information from multiple sources.” (Idem.)

Continued on next page

NCTM Assessment Perspectives

Continued from previous page

NCTM 1995 Assessment Standards for School Mathematics

The Assessment Standards indicate that exemplary mathematics assessment should:

- Reflect the mathematics that students should know and be able to do
- Enhance mathematics learning
- Promote equity
- Be an open process
- Promote valid inference

National Assessment Programs

National Adult Literacy Survey (NALS)

A national survey reported in 1993, which provided a profile of the literacy skills of the United States' adult population. The results revealed that more than 40% of all American adults have literacy levels at Levels 1 or 2 (out of five), below the level required to secure jobs at good wages.

National Reporting System (NRS)

An outcome-based reporting system for the state-administered, federally funded adult education program required by Title II of the Workforce Investment Act. The goals of the NRS were to establish a national accountability system for education programs by identifying measures for national reporting and their definitions, establishing methods for data collection, developing software standards for reporting to the U.S. Department of Education, and developing training materials and activities on NRS requirements and procedures.

Although this may seem a paradox, all exact science is dominated by the idea of approximation.

—Bertrand Russell (1872-1970)

The Math Bulletin is a publication of SABES, the System for Adult Basic Education Support, and is funded by the Massachusetts Department of Education.

