

Augustana College

EDUC 245 – Educational Psychology & Measurement

Concepts addressed:
Assessment Strategies

Standardized Testing - These tests are usually given nationwide under uniform conditions (set amount of time, standard directions) and are corrected in an objective manner. There are three types of standardized tests, achievement, diagnostic and aptitude.

1. Achievement tests assess the amount students learned in a certain subject such as reading or math (Iowa Basic Skills Test).
2. Diagnostic tests identify students' weaknesses and strengths; these are most commonly used to determine special education needs. Usually a school psychologist or special education teacher administers them to individual students.
3. Aptitude tests measure abilities over a long time period and are used to predict future success (IQ, ACT).

Traditional Testing - Objective tests and essay tests are the two most prevalent types of traditional testing. Objective tests include multiple-choice, true/false, fill-in the blank and matching items. This method of testing is helpful when the teacher wants to cover a large amount of information, but not go into great depth. Objective tests primarily assess knowledge and comprehension (Bloom's Taxonomy) of information. Multiple-choice items can be constructed to tap into higher-level thinking (application, analysis, synthesis, evaluation). Essay tests are beneficial when tapping into higher level thinking, they are designed to cover a smaller portion of content.

Performance based (authentic tests) require students to apply content to real-life situations. This type of assessment happens over a period of time (not a 50 minute test), involves some collaboration, taps into higher level thinking skills, allows for differing learning styles, interests, and abilities, and usually involves an audience. Examples are developmental portfolios (collection of student's work over time); showcase portfolios (best work), exhibitions, and other real-life applications (playing a musical instrument, speaking a foreign language, debating, simulation, role play).

Informal assessments can easily be incorporated into classroom routines and learning activities. They are indicative of the student's performance on the skill or subject of interest. Examples of informal assessments include:

- Checklists/observations
- Anecdotal notes
- Running records
- Student portfolios
- Teacher/student conference notes
- Learning logs

Norm referenced tests are used to make comparisons between groups. They are useful in selecting a few top students for a program and in assessing overall achievement. They are not useful in finding out if a group is ready to progress to the next level.

Criterion-referenced tests compare student achievement to a standard of performance. They measure the mastery of specific objectives. They can be used to determine students' strengths and weaknesses.

Summative and Formative Assessment - two different purposes for testing

Formative assessment - testing that is used to inform instruction. It may be done before material is presented (pre-test) or during instruction. It reveals the students' current understanding of the material and any areas of weakness. Formative tests are non-graded.

Summative assessment - occurs at the end of instruction to assess achievement and is graded. The same test may be used for formative and summative assessment, but the purpose of the assessment is different.

Validity and Reliability - all assessment must have validity and reliability if the results are to be interpreted accurately.

Validity ensures the assessment instrument measures what it is suppose to be measuring. The method of assessment needs to fit the learning outcomes and the emphasis on the objectives should parallel the lessons taught. For example, if the purpose of the test is to determine reading comprehension the assessment should reflect the student understanding of the reading material.

Reliability means the assessment is replicable by others and the results about the student's knowledge, skills or abilities are consistent in different circumstances. Standardized tests and teacher made tests ensure this in the following manner:

1. Standard time for the test
2. Written set of instructions for the teacher
3. Assessment aligns with the learning objectives
4. Uncomplicated language
5. Enough questions to cover the content
6. If using objective tests one clear, right answer.

Standardization - teachers give similar assignments, instructions, time frame and resources for assessment to all students (this may be tempered by student ability and interest).

Rubrics: on a point scale differentiates between levels of performance and specifies the type of performance that is expected for each rating (6 Trait Writing).

Checklists: list of evaluation criteria, the student has either completed the activity or not. Student and teacher may sign off on a checklist (list items to complete in a research project).

Rating scale: teacher or student chooses a response from a variety of categories. The categories are usually in hierarchical order.

Scoring essay tests: construct a model answer, weight areas according to the degree of importance, ensure anonymity of students while correcting (reduces the subjectivity), score all responses to one question before proceeding to the next question, if reading multiple essays mix up the papers between questions (counters teacher exhaustion bias), and don't over penalize spelling and handwriting errors. Rubrics are often helpful in scoring essay tests.

Scoring objective tests (multiple-choice, true/false, fill in the blank, matching): there is one correct answer. Some performance assessment can be done objectively (timed test in math, word processing test, 50 meter dash).

The primary purpose of assessment is to aid teachers in making good decisions that will foster learning in academic subjects. Assessment aids teachers in making decisions about the individual student, the curriculum content and instructional strategies. Specifically, assessment is used to:

1. Determine student grades and to report student progress to students, parents, the principal, and other teachers
2. Hold schools accountable to the public (No Child Left Behind).
3. Assign students to certain groups and to determine if they should be promoted or
4. Graduated (high-stakes testing).
5. Determine special education needs for students.
6. Place students in gifted and talented programs.
7. Admit students into college.
8. Determine readiness (kindergarten readiness or next level of education).
9. Motivate students to increase effort (students and their families' value good grades).
10. Certification of teachers.

Measurement is evaluation that is put into quantitative (number) form.

Frequency Distributions are a distribution (any set of scores) that is tallied according to the number of people that received each score.

Measures of Central Tendency include mean (arithmetic average), median, and mode. Usually mean is considered the most useful. A normal distribution is formed when the mean, median, and mode are the same number.

Standard Deviation (sd) is how the scores vary from the mean (variability). 1 sd is always where 68% of the scores are located. If the mean is 10 and the sd is 2 then 68% of the scores fall between 8 and 12. The smaller the sd the more homogeneous the class.

Percentile Rank Scores reflect the percentage that scored at or below an individual's score. If a student is in the 70th percentile it means their raw score was the same or better than 70% of the students that took the test. Percentile is easily communicated to parents.

Grade-Equivalent Scores are reported in terms of a particular grade level that is most similar to the score the test taker received on a test. A 4th grader excelling in reading may have a 9th grade equivalent score. This form of reporting is confusing and misleading to parents.

Stanine Scores are standard scores that compare student's rank. Scores range from 1 (very low) to 9 (extremely high). They are always reported in whole numbers. Each stanine score represents a

Development of this review sheet was made possible by funding from the US Department of Education through South Dakota's *EveryTeacher* Teacher Quality Enhancement grant.

wide range of raw scores enabling student's score to be viewed in more general terms. They are easily communicated to parents.

z-scores are standard scores with a mean of 0 and a standard deviation of 1. z-scores are reported on a continuum from -3 to +3 (z-score of +2.3 would indicate high achievement).

Confidence Interval - Range of scores that an individual may score on a particular test. If the confidence interval is 5 on a test and a student receives a 72 the teacher would assume that on any given day the student would score between a 67 and a 77.

Correlation - shows a relationship between two or more variables (it does not show cause and effect). It is reported on a + 1 to -1 continuum. + 1 indicates a strong positive correlation (both variables score high - academic achievement and salary), 0 denotes no correlation (eating pie and social skills), and -1 shows a strong negative correlation (one score is high, the other is low - sunny day and the number of umbrellas).

Multiple assessments (formal and informal) are needed to determine a grade that accurately reflects student performance. Grades may include scores on quizzes, tests, homework, participation, and effort. Grades are usually reported quarterly (every 9 weeks). Feedback should be continual throughout the school year, so grades are never a surprise to a student. Letter grades should be accompanied with some sort of written or verbal communication. Letter grades are not developmentally appropriate at the primary level. Grades should be kept confidential.

Criterion-referenced grading: students are graded according to their performance on certain criteria. The report card lists the objectives and the teacher indicates the level of student proficiency. Student's level of proficiency cannot be averaged into a single grade.

Norm-referenced grading: students are graded in comparison to their classmates. The report card merges all assignments and tests into one score.

Point system: every test and assignment is given a certain number of points according to difficulty and importance. To prepare a final grade points are added up and then grades are assigned by percentage (90% and above is an A) or by looking for natural gaps (99, 99,98,92,91 - the gap between 98 and 92 would divide the A and B grades).

Percentage Grading: every test and assignment is graded on the percentage correct (A = 90 to 100% correct).

Contract Grading: the student works for a particular grade according to work designated on a contract. Teachers must be sure to ensure quality and not just quantity when designing the contract. This type of grading usually reduces anxiety.

Grading on effort: elementary teachers may have a dual marking system, one grade indicates academic achievement and the other indicates effort in relation to ability.

Conferences are important at both the elementary and secondary levels. They are an opportunity for direct communication between the teacher and family. In order to hold a successful conference it is important for teachers to be well prepared (grade book, examples of work, as

Development of this review sheet was made possible by funding from the US Department of Education through South Dakota's *EveryTeacher* Teacher Quality Enhancement grant.

much objective data as possible). Begin and end with a positive statement. Identify the goals for the conference (academic achievement, behavior). Be specific about concerns and have suggestions for working with the parents. Listen carefully when the parent or guardian speaks. Document important information covered at the conference. If appropriate, set up additional conference times.