CHAPTER X: Assessing Students' Learning

1. DEFINITION

Evaluation, or assessment, refers to all the means used in schools to formally measure student performance (McMillan, 2004; Popham, 2005). These include quizzes and tests, written evaluations, and grades. Student evaluation usually focuses on academic achievement, but many schools also assess behaviors and attitudes.

Tests and grades tell teachers, students, and parents how students are doing in school. Teachers can use tests to determine whether their instruction was effective and to find out which students need additional help. Students can use tests to find out whether their studying strategies are paying off. Parents need grades to learn how their children are doing in school; grades usually serve as the one consistent form of communication between school and home.

2. ASSESSMENT OBJECTIVES

Student evaluations serve six primary purposes (see Gronlund, 2003):

2.1. Feedback to students

Teachers and students need to low as soon as possible whether their investments of time and energy in a given activity are paying off by the increasing of their learning. For example, suppose a teacher had students write compositions and then gave back written evaluations. Some students might find out that they needed to work more on content, others on the use of modifiers, still others on language mechanics. This information would help students to improve their writing much more than would a grade with no explanation.

2.2. Feedback to teachers

One of the most important (and often overlooked) functions of evaluating student learning is to provide feedback to teachers on the effectiveness of their instruction. Teachers cannot expect to be optimally effective if they do not know whether students have grasped the main points of their lessons. Asking questions in class and observing students as they work gives the teacher some idea of how well students have learned; but in many subjects brief, frequent quizzes, writing assignments, and other student products are necessary to provide more detailed indications of students' progress. Evaluations also give information to the principal and the school as a whole, which can be used to guide overall reform efforts by identifying where schools or subgroups within schools are in need of improvement (Hanna & Dettmer, 2004)

2.3. Evaluation as Information

A report card is called a report card because it reports information on student progress. This reporting function of evaluation is important for several reasons.

2.4. Information to parents

First, routine school evaluation of many kinds (tests, and certificates as well as report card grades) keep parents informed about children's schoolwork. For example, if a student's grades are dropping, the parents might know why and might be able to help the student get back on

track. Second grades and other evaluations set up informal home-based reinforcement systems.

2.5. Information for selection and certification

Some sociologists see the sorting of students into societal roles as a primary purpose of schools. Closely related to selection is certification, a use of tests to qualify students for promotion or for access to various occupations. For example, many states and local districts have minimum competency tests that students must pass to advance from grade to grade or to graduate from high school.

2.6. Information for accountability

Often, evaluations of stdents serve as data for the evaluation of teachers, schools.

2.7. Evaluation as Incentive

One important use of evaluations is to motivate students to give their best efforts. In essence, high grades, stars, and prizes are given as rewards for good work. Students value grades and prizes primarily because their parents value them.

3. HOW IS STUDENT LEARNING EVALUATED ?

To understand how assessments can be used most effectively in classroom instruction, it is important to laow the differences between formative and summative evaluation and between norm-referenced and criterion-referenced interpretation.

3.1. Formative and Summative Evaluations

A *formative evaluation* asks, "How well are you doing and how can you be doing better?" A *summative evaluation* asks, HOW well dld you do?" Formative, or diagnostic, tests are given to discover strengths and weakessesin learning and to malte midcourse corrections in pace or content of instruction. Formative evaluations might even be made "on the fly" during instruction through oral or brief written learning probes.

  Formative evaluation is useful to the degree that it is informative, closely tied to the curriculum being taught, timely, and frequent (McMillan, 2004). For example, frequent quizzes that are given scored immediately after specific lessons might serve as formative evaluations, providng feedback to help both teachers and students improve students'learning.

In contrast, *summative evaluation* refers to tests of student laowledge at the end of instructional units (such as final exams). Summative evaluations may or may not be frequent, but they must be reliable and (in general) should allow for comparisons among students. Summative evaluations should also be closely tied to formative evaluations and to course objectives.

Some Examples of Summative and Formative Assessment

Formative	Summative
• In-class discussions • 1 on 1 conversations with students • Peer-generated concept maps • Exit tickets written at the end of class • Stoplight check in of their understandings • Peer-review of assignments • In Class Quiz (not contributing to the overall course grade)	• Midterm or Final Exam The following are examples provided they contribute to the overall course grade: • Online or In-Class Quiz • Paper • In-Class Debate • Final Recital or Performance

4. NORM-REFERENCED AND CRITERION-REFERENCED EVALUATIONS

The distinction between norm-referencing and criterion-referencing refers to how students' scores are interpreted.

4.1. Norm-referenced interpretations

Focus on comparisons of a student's scores with those of other students. Within a classroom, for example, grades commonly are used to give teachers an idea of how well a student has performed in comparison with classmatesA. student might also have a grade-level or school rank; and in standardized testing, student scores might be compared with those of a nationally representative norm group.

4.2. Criterion-referenced interpretations

Focus on assessing students' mastery of dpecific slulls, regardless of how other students did on the same slulls. Criterionreferenced evaluations are best if they are closely tied to specific objectives or well specified domains of the curriculum being taught. Comparison of the principal features and purposes of criterion-referenced and norm-referenced testing (Popham, 2005).

Table : Comparison of the principal features and purposes of criterion-referenced and norm-referenced testing (Popham, 2005).

Feature	Norm-Referenced Testing	Criterion-Referenced Testing
Principal use	Survey testing	Mastery testing
Major emphasis	Measures individual differences in achievement	Describes tasks students can perform
Interpretation of results	Compares performance to that of other individuals	Compares performance to a clearly specified achievement domain
Content coverage	Typically covers a broad area of achievement	Typically focuses on a limited set of learning tasks
Nature of test plan	Table of specifications is commonly used	Detailed domain specifications are favored
Item selection procedures	Items selected to provide maximum discrimination among individuals (to obtain high score variability); easy items typically eliminated from the test	Includes all items needed to adequately describe performance; no attempt is made to alter item difficulty or to eliminate easy items to increase score variability
Performance standards	Level of performance determined by *relative* position in some known group (e.g., student ranks fifth in a group of 20)	Level of performance commonly determined by *absolute* standards (e.g., student demonstrates mastery by defining 90 percent of the technical terms)

Source: Adapted from Norman E. Gronlund, *How to Make Achievement Tests and Assessments* (5th ed.). Copyright © 1993 by Allyn & Bacon. Reprinted by permission.

5. MATCHING EVALUATION STRATEGIES WITH GOALS

Teachers must choose different types of evaluation for different purposes. At a minimum, two types of evaluation should be used: one directed at providing incentive and feedback and the other directed at ranking individual students relative to the larger group.

5.1 Evaluation for Incentive and Feed back

It is important to have a clear and objective set of criteria that student work is compared with so students can see exactly why they scored as they did. If the criteria are illustrated using a rubric that has descriptions of different levels of achievement (scores) as well as examples of student wok at the highest levels of achievement (or better yet, that is typical of each possible score students might receive according to the rubric), then students can see exactly how their achievement compares with the criteria.

5.2 Evaluation for Comparison with Others

There are times when teachers need to know and to communicate how well students are doing in comparison to others. This information is important to give parents (and students themselves) a realistic picture of student performance.

Comparative evaluations are traditionally provided by grades and by standardized tests. Unlike incentive/feedbaclc evaluations, comparative evaluations need not be conducted frequently. Rather, the emphasis in comparative evaluations must be on fair, unbiased, reliable assessment of student performance. Comparative evaluation should sess what students can do and nothing else.

6. HOW ARE TESTS CONSTRUCTED?

Writing good achievement tests is therefore a critical skull for effective teaching. This section presents some basic principles of achievement testing and practical tools for test construction.

7. PRINCIPLES OF ACHIEVEMENT TESTING

Gronlund (2000) listed six principles to keep in mind in preparing achievement tests. These are paraphrased as follows:

a. Achievement tests should measure clearly defined learning objectives that are in harmony with instructional objectives.

b. Achievement tests should measure a representative sample of the learning

tasks included in the instruction.

c. Achievement tests should include the types of test items that are most appropriate for measuring the desired learning outcomes. Items on achievement tests should correspond as closely as possible to the ultimate instructional objectives.

d. Achievement tests should fit the particular uses that will be made of the results. Each type of achievement test has its own requirements. For example, a test that is used for diagnosis would focus on particular slzills with which students might need help.

e. Achievement tests should be as reliable as possible and should be interpreted with caution. A test is reliable to the degree that students who were tested a second time would fall in the same rank order. In general, writers of achievement tests increase reliability by using relatively large numbers of items and by using few items that almost all students get right or that almost all students miss.

f. Achievement tests should improve learning. Achievement tests of all kinds, particularly formative tests, provide important information on students' learning progress. Stiggins (2004), for example, urges that assessments *for* learning are more important than assessments *of* learning. Achievement testing should be seen as part of

the instructional process and should be used to improve instruction and guide student

Learning.

8. TYPES OF LANGUAGE TESTS

8.1   Achievement Test

An achievement test, also referred to as attainment or summative test, are devised to measure how much of a language someone has learned with reference to a particular course of study or programme of instruction, e.g. end-of-year tests designed to show mastery of a language. An achievement test might be a listening comprehension test based on a particular set of situational dialogues in a textbook. The test has a two-fold objective: 1) To help the teachers judge the success of their teaching. 2) To identify the weaknesses of their learners. In more practical and pedagogical terms, Brown (1994, p. 259) defines an achievement test as „tests that are limited to particular material covered in a curriculum within a particular time frame". In other words, theyare designed primarily to measure individual progress rather than

as a means of motivating or reinforcing language. Ideally, achievement tests are rarely constructed by classroom teacher for a particular class.

8.2   Diagnostic Test

As its name denotes, a diagnostic test is primarily designed to diagnose some particular linguistic aspects. Diagnostic tests in pronunciation, for example, might have the purpose of determining which particular phonological features of the English language are more likely to pose problems and difficulties for a group of learners. One of the well-known diagnostic tests in English is Prator"s (1972) Diagnostic Passage. It consists of a short written passage that the learner reads orally; the teacher then examines a tape recording of that reading against a very detailed checklist of pronunciation errors. Basically, diagnostic language tests have a three-fold objective:1.Toprovide learners with a way to start learning with their own personal learning programme or what would be called in the literature of testing learning paths.2.To provide learners with a way to test their knowledge of a language.3.To provide learners with better information about their strengths and weaknesses. Ideally, diagnostic tests are designed to assess students" linguistic knowledge (knowledge of and about the language) and language skills (listening, speaking, reading and writing) before a course is begun. However, the termformativeis sometimes used to designate a diagnostic test. One of the main advantages of a diagnostic test is that it offers useful pedagogical solutions for mixed-ability classes. In this very specific context, Broughton et al. (1980) contend that: There will certainly be a large block in the middle of the ability range who can be separated off as a group for some parts of the lesson, or for some lessons, and will form a more homogenous teaching group. If this strategy is adopted, the poor ones and the better ones must receive their due time and attention. (Broughton et al. 1980, p. 189)

8.3   Language Aptitude Test

Before one ventures into defining what a language aptitude test is, it would be wiser to start first by defining what a language aptitude is. Language aptitude, as a hybrid concept part linguistic and part psychological, refers to the genuineability one is endowed with to learn a language. It is thought to be a combination of several abilities:

- **Phonological ability,i**.e. the ability to detect phonetic differences (e.g. of stress, intonation, vowel quality) in a new language.

- **Syntactic ability**, i.e., the ability to recognize the different grammatical functions of words in sentences.

- **Psychological ability**, i.e. rote-learning abilities and the ability to make inferences and inductive learning.

Additionally, Crystal (1989, p. 371) suggests other variables conducive to successful language learning such as empathy and adaptability, assertiveness and independence with good drive and powers of application". A high language-aptitude person can learn more quickly and easily than a low language-aptitudeindividual. The evidence in such assertion is axiomatic in a language aptitude test. A language aptitude test tends to measure a learner aptitude for language learning, be it second or foreign, i.e. students performance in a language. Thus, it is usedto identify those learners who are most likely to succeed. Language aptitude tests usually consist of several different test items which measures such abilities as:

- **Sound-coding ability,** i.e. the ability to identify and remember new sounds in a new language.

- **Grammar-coding ability**, i.e. the ability to identify the grammatical functions of different parts of sentences.

- **Inductive-learning ability**, i.e. the ability to work out meanings without explanation in the new language.

- **Memorization**, i.e. the ability to remember and to recall words, patterns, rules in the new language

8.4   Placement Test

A placement test, as its name implies, is originally designed to place learners at an appropriate level in a programme or course. The term "placement test" as Richards et al. (1989) note does not refer to what a test contains or how it is constructed, butto the purpose for which it used. Various types or testing procedures such as dictation, interview or a grammar test (discrete or integrative) can be used for placement purposes. The English Placement test (EPT), which is a well-known test in America, is an illustrative example of this

test-type. The EPT is designed to assess the level of reading and writing skills of entering undergraduate students so that they can be placed in appropriate courses. Those undergraduate students who do not demonstrate college or university-level skills will be directed to remedial courses or programmes to help them attain these skills.

8.5 Proficiency Test

A proficiency test is devised to measure how much of a language someone has learned. It is not linked to any particular course of instruction, but measures the learner"s general level of language mastery. Most English language proficiency tests base their testing items on high frequency-count vocabulary and general basic grammar. Some proficiency tests have been standardized for worldwide use, such as the well-known American tests, the TOEFL, and the English Language Proficiency Test (ELPT)3which are used to measure the English language proficiency of foreign students intending further study at English-speaking institutions, namely the USA. However, the Cambridge Certificate of Proficiency in English or CPE, as it is generally referred to, is the most advanced remains the only British top-value and high-prestige standardized4language test. It is the most advanced generalEnglish exam provided by the University of Cambridge. The Certificate is recognized by universities and employees throughout the world. The English level of those who pass the CPE is supposed to similar to that of a fairly educated native speaker of English. Clearly, as Valette posits, „the aim of a proficiency test is to determine whether this language ability corresponds to specific language requirements"(Valette, 1977, p. 6) Actually, there are four other types of Cambridge proficiency tests, the Cambridge Key English Test (KET), the Cambridge Preliminary English Test (PET), The Cambridge First Certificate of English (FCE) and the Cambridge Certificate in Advanced English (CAE). The material contained in proficiency tests can be used for teaching as well as for testing. In essence, a proficiency test measures what the student has learned in relation to a specific purpose, e.g. does the student know enough English to follow a course offered in English?

8.6 Progress Test

A progress test is an achievement-like test. It is closely related to a particular set of teaching materials or a particular course of instruction. Progress tests are usually administered at the end of a unit, a course, or term. A progress test may be viewed as similar to an achievement

test but much narrower and much more specific in scope (Richards et al., 1989). They help examiners in general and language teachers in particular to assess the degree of success of their programmes and teaching and therefore to identify their shortcomings and weaknesses respectively. Progress tests can also be diagnostic to some degree, in the sense that they help identify areas of difficulties encountered by learners in general.