Role of Classroom Testing

Testing has evolved and has become extremely prevalent in our educational system today. Year after year, testing is increasingly becoming an issue of concern and as previously mentioned, classroom tests have their share in the language/teaching process. Yet the grading function should not be overemphasized at the expense the learning function. Valette (1977) notes that in-class testing fulfills three main functions in second and foreign language learning. These functions can be summed up as follows:

1. Definition of course objectives.
2. Stimulation of student progress.
3. They evaluate classroom achievement.

Definition of Course Objectives

From an instructional standpoint, classroom tests are used in a very helpful way to define the course objectives. In other words, they define the short-term course objectives envisioned by the teacher, as well as the content and nature of the language learning programme. This has a two-fold aim, for one thing, the teacher will be systematically geared towards the set objectives; for the other, the tests results will indicate how close the learner has come to attaining the objectives. In sum as Haertel (1999) posits, testing appears to be the logical approach to identify learners who do not meet expectations.

Stimulation of Student Progress

Traditionally tests have been devised by teachers to point up the learner's ignorance, errors and lack of application. However, from a didactic perspective, testing is supposed to offer ample opportunities for the learner to measure how well he masters specific linguistic items of the target language. In this context Valette argues that *'The test best fulfills its function as part of the learning process if correction performance is immediately confirmed and errors are pointed out'* (Valette, 1977, p. 4).

Evaluation of Classroom Achievement

Most teachers contend that testing is all too often viewed as a necessary evil. On should also note that testing on a frequent and regular basis provides the teacher with valuable information concerning areas of difficulties the learners encounter. In this way, the teacher gets more about what aspects of language need further clarification and explanation and subsequently devise remedial activities. What is more, testing provides the teacher with clues and details related to the effectiveness and soundness of a specific teaching approach and method. It also gives an objective evaluation of learner's progress individually, his attainment of the set course objectives and his performance in relation to that of the other classmates. In sum, one should view testing as a bridge-building process between teaching and learning and classroom tests as mirrors in which teachers and students see their reflections clearly (Valette 1977).

Definition of Test

In educational terms, a test can be defined as '***any procedure for measuring ability, knowledge and performance'*** (Richards et al.*,* 1985, p. 291), while Brown (1994, p. 252) notes that a test is ***'a method of measuring a person's ability or knowledge in a given area'.*** In very practical terms, tests yield scores that mirror attributes or charcteristics of individuals (Allan, 1995). Brown's definition seems to be more comprehensive in the sense that it covers all the main components of a test. However, what does a test consist of and what is it intended for? Firstly, a test is a *method* consisting of a set of techniques, procedures and test items that constitute an instrument of some sort. Secondly, a test has the purpose of *measuring* the testee's performance in precise mathematical terms, assigning a grade, or expressing evaluative qualifiers, such as excellent, good, fair, poor and so on. Thirdly, a test is intended to measure a *person's* ability or knowledge, i.e. who are the testees and what is, for example, their linguistic background knowledge? Next, a test measures the *ability* or *knowledge*, that is to say, competence and know-how. Finally, the test is closely related to a *given area*, in the case of a proficiency test, that area is language proficiency, e.g. communicative competence. In other terms, ***"The overall purpose of any form of language testing is to sample the language abilities of candidates in such a way that a realistic representation of their degree of skill in using language in non-test situations is provided."*** (Milanovic, 2002, p. 2).

Devising a language test that accounts for the different linguistic variables is not an easy task. Broughton et al. (1980) note that the preparation of tests for educational measurement is time-consuming, expensive and requires expertise in statistical techniques as well as in devising suitable tasks for the linguistic assessment to be based on. Additionally, Brown (1994) remarks that, *'one of the biggest obstacles to overcome in constructing adequate tests is to measure the criterion and not inadvertently something else'* (Brown, 1994, p. 253). In this sense, he puts forward three requirements that are axiomatic to qualify a test as a ''good'' test: practicality, reliability and validity; in Brown's view, if these three requirements are carefully met, a test can be administrable, dependable and can actually measure what it intends to measure. On the other hand, Flavell (1983) notes that a teacher who ignores the interrelatedness between the content of a test and the consistency of the results it gives is in danger of writing tests which are likely to produce misleading information about the test-takers.

1. Practicality

It is highly recommended that some practical considerations are to be taken into when constructing and administrating a test. These considerations closely relate to financial means, time constraints, ease of administration, and scoring and interpretation. Undoubtedly, a test which requires considerable financial means and therefore a considerable budget is impractical. Additionally, a test which is time-consuming in the sense that uses up hours and hours to complete is also impractical. Finally, a test which requires individual one-to-one testing is impractical for hundreds of people and only a limited number of examiners. Conversely, a test that takes a few minutes for a student to complete and several hours for the examiner to correct is impractical. A test that is too complex and too sophisticated may not be of practical use to the teacher. In other words, it lacks *instructional value* to use Oller's (1979, p. 52) terms. In sum, the value, quality, credibility and formality of a test are largely dependent upon such basic facts and realities.

2. Reliability

The criterion of reliability in test constructing denotes the degree to which a test gives consistent results. Actually, a test is said to be reliable if it gives the same results repeatedly when it is given on different occasions, or it is used by different people. Generally, if people get similar scores on parallel forms of a test, i.e. using different forms of a test which try to measure the same skills and abilities using the same methods of testing, equal length and level

of difficulty, this proves that the test is reliable. Harmer (2001) notes that, "*In practice, 'reliability' is enhanced by making the test instructions absolutely clear, restricting the scope for variety in the answers, and making sure that the test conditions remain constant*" (Harmer, 2001, p. 322). However, it is worth noting that the careful specification of an analytical scoring instrument can increase, what Brown and Bailey (1984) have called *the scorer reliability*, to refer to the consistency of scoring by two or more scorers or examiners. Put differently, the circumstances in which the test is taken, the way in which it is marked and the uniformity of the assessment it makes (Flavell, 1983). In sum then, for Lado (1961) reliability is seen as a prerequisite for validity (Xi, 2008).

3. Validity

Arguably, the attribute of validity is in effect complex and multi-faceted. Basically, it refers to the degree to which a test measures what is supposed to be measured, or can be used successfully for the purposes for which it is intended. In other words, does the test evaluate what is intended to evaluate? For example, Harmer (2001) notes that, *"to test writing ability with an essay question that requires specialist knowledge of history or biology- unless it is known that all students share this knowledge before they do the test."* Teachers can use a set of different statistical procedures to apply to a test to evaluate its validity. Such procedures seek to determine what the test actually measures, and how to what extent it does so. However, a question is worth posing this level: how are teachers to establish the validity of a test? The answer to this question leads us to explore other related aspects of validity; the following are of capital importance for the classroom teacher: content validity, construct validity, empirical validity, and face validity. In gross, teachers, all insist, in somewhat different ways, that test validity must account of how and where a test is used.

3.1. Content Validity

This aspect of validity is based on the degree to which a test adequately and sufficiently measures the particular skills it sets out to measure, what is called content specification, in other words, the extent to which the content of the test matches the instructional objectives. For example, a test of pronunciation skills in a language learning programme would have low validity content if it tested only some of the skills which required accurate pronunciation, such as a test which tested the ability to pronounce isolated words with no reference to the other supra-segmental phonological features as stress, intonation and pitch. In this very specific context Flavell posits that **"*The content specification is important because it ensures as far***

as possible that the test reflects all the areas to be tested in suitable proportions and also because it represents a balanced sample, without bias towards the test material which happens to be available" (Flavell, 1983, p. 11)**.** A related point worth raising here is that content validity is crucial for the teacher who sets his own tests. However, according to Lado (1961), content validity concerns the degree to which an item contains a language problem that is representative of the problem in real life. As Xi (2008, p. 178) posits, *"A direct language test has to show face or content validity by demonstrating its resemblance of 'real-life' language situations in the setting and linguistic content."*

3.2. Construct Validity

The late 1970s and early 1980s saw the first hint of the notion of construct validity in language testing (Xi, 2008). This aspect is based on the degree to which the items in the test reflect the theory or the construct on which the test is based. For example, in language proficiency the greater the relationship which can be demonstrated between a test of communicative competence in a language and the theory relating to this concept, the greater the construct validity of the test.

3.3. Empirical Validity

This aspect measures the validity of a test arrived at by comparing the test with one or more criterion measures, i.e. another or other tests which are known to be valid. Such comparison could be made on the following basis:

1. Other valid tests or other independent measures obtained at the same time, e.g. an assessment made by the teacher.

2. Other valid test or other performance criteria obtained at a later time.

3.4. Face Validity

The criterion of face validity refers to the degree to which a test appears to measure the knowledge or abilities it claims to measure, making use of the subjective judgment of an observer. Put differently, and to use Brown's (1994, p. 256) question: *"does the test, on the face of it, appear to measure what it is designed to test?"* For example, if a test related to a reading comprehension lesson or course contains many dialect or slang words which the students are very likely to ignore, the test may be said to systematically lack face validity. Additionally, one way of finding out more about the notion of face validity is simply to ask

teachers and students concerned for their opinions and views about the test. This could be done either formally by administrating a questionnaire or through an in-class informal discussion.

Principles of Testing

Many teachers still hold a specific vision about testing. They all too often regard it as one of the most controversial areas of the teaching/learning process. It is undeniably an in-class activity that is necessary as a form of completion of the teaching input and the learning output. Basically, if properly prepared and adequately implemented, testing undoubtedly turns to be an objective pedagogical tool serving as activity to check the effectiveness of the whole language teaching/learning process. Test scores provide a valuable measure of how well the curriculum is being learnt and help indicate how well students do at the main exit points of the school system, for example the baccalaureate exam. To fulfill faithfully the functions that are assigned to testing, teachers should turn their attention towards the following basic principles of testing:

1. To assess learners' performance in the target language the teacher should not give a task that the learner cannot perform. The task should be *authentic*, *realistic* and *appropriate* to their linguistic level.

2. Even when assessing the learners' performance, at any level, the learners should be given clear instructions well. They should know what they are expected to do in a given task. The ideas, feelings and emotions that the learners want to express cannot be limited to their insufficient linguistic input.

3. Teachers should test the outcomes or products of what they have taught their learners, not what their colleagues know.

4. Teachers should not use a technique not used in the teaching process as a test technique to have a positive washback affect of testing on language learning and teaching (see 4.5.).

5. Teachers should test learners' writing skills by having them write and their speaking by having them speak. This is what is known as 'construct validity (see 2.5.3.2.).

6. We teach people and we evaluate language ability but we do not evaluate people.

(Adapted from Korsal, 2006)