

CHAPTER THREE

CHARACTERISTICS OF ASSESSMENT

Assessment, if it is to adequately fulfill its purpose, should meet the three requirements. The test which is appropriate in terms of our objective is a **valid** test. If the evidence it provides is dependable, it is a **reliable** test. Also, if it can do this with ease and economy, it is a **practical** test. We shall now examine the qualities of **validity**, **reliability**, and **practicality** in detail.

3.1 VALIDITY

Validity is the degree to which a test measure what it is supposed to measure or can be used successfully for the purpose for which it is intended. Two questions must always be considered:

1. What precisely does the test measure?
2. How well does it do it ?

Types of validity

(a) Content Validity

Almost certainly the most important for the practicing teacher is the extent to which *a test adequately covers the syllabus to be tested*. A valid test must be based upon a careful analysis of the subject or skill we are testing. It must also be constructed as to represent adequately each portion of this analysis. If, for example, a test is to measure students' mastery of English grammatical structures, the teacher has to study the unit and pick out the new structure points. He has then to plan how the students are to demonstrate mastery of these structures in behavioural terms. This is achieved by the use of a specification table of content and behaviour (i.e., recognition or production).

(b) Empirical Validity

If we are to check the effectiveness of a test and to determine how well the test measures, we should relate the test scores to some independent, outside criteria such as scores given at the end of the course or the teacher's judgment of his students or an external examination.

The relationship between test scores and students' performance in other measures

(c) Face Validity

This simply means the way the test looks to the examinees, supervisors or in general to the people concerned with the education of the students. Obviously face validity can never replace content validity or empirical validity; yet its importance should not be underestimated.

The degree to which an assessment or a test appears to measure what is supposed to measure.

(d) Construct Validity

A test is said to possess construct validity if *it is designed in accordance with a theory of language behaviour and learning*. For example, if the course of study is based on the assumption that systematic language habits are best acquired by means of the structural approach then a test, which emphasizes the situational meaning of language, will have low construct validity.

3.2 RELIABILITY

Reliability means the stability (consistency) of test scores. Probably, if the same test is given twice to the same group of students, under the same conditions, it would give the same results. The requirements of a dependable test are the following:

(1) Multiple samples

The more samples of students' performance we take the more reliable will be our assessment of their knowledge and ability. Consequently, the test must be long enough to provide a generous sampling of the areas tested. It should also contain a wide variety of levels of difficulty.

(2) Standard Conditions

The reliability of the test scores can be assured only if all students take the examination under identical conditions. In a listening test, for example, all students must be able to hear the items clearly.

(3) Standard Tasks

All students must be given the same items or items of equal difficulty. In other words, the tests must be identical so as the format.

(4) Standard Scoring

All test must be scored in an identical manner. The scorer should give the same or nearly the same score repeatedly for the same test performance. Two or more scorers should give equivalent score for the same performance. Objective tests tend to be more reliable than free response tests like composition, where *individual judgement* must be made.

3.3 PRACTICALITY

A third characteristic of assessment is its practicality or usability. A test which is valid and reliable but difficult to administer or score or which is expensive may fail to gain acceptance.

Two parameters appear to be involved:

(1) Economy: The cost in time, money and personnel of administering a particular test .

(2) Ease: The degree of difficulty experienced in the administering and scoring of the test, for example, an oral test that demands the use of a tape recorder is not practical if it has to be administered to thousands of students.