CHAPTER 3

ASPECTS OF FORMATIVE ASSESSMENT

3.0. Introduction

This chapter mainly aims to discuss ‘Assessment’ in general and ‘Formative Assessment’ in particular. However, in order to give a clear perspective to the concept of ‘Assessment/Formative Assessment’, an attempt has been made to first define and discuss the features of other related terms like ‘testing’ and ‘evaluation’, which are often used interchangeably by the stakeholders.

3.1 Language Testing

‘Tests/Testing’ is the earliest and most commonly used term, which is used right from the very beginning to make a judgment of the learner’s achievement at the end of an educational program. This sort of practice was included in the traditional testing in which the only purpose is to weigh the learner’s linguistic competence without taking into consideration the effect of tests or wash back effect on the students. Testing can be described as a “device” or “instrument”, which measures the linguistic knowledge or competence of the learners. In other words, it has been defined as a “measuring device” for assigning numerical value to the response for a given task to each of the members or a set of objects or group of persons, normally examinees.

*Longman Dictionary of Applied Linguistics*, edited by Richards, Platt and Weber, defines *tests* as “any procedure for measuring ability, knowledge, or performance” (p.291). Thus, this definition considers it as a procedure and appears to limit the term ‘testing’ mainly to learners.

Ingram (1974, p.313) claims:

Tests, like examinations, invite candidates to display their knowledge or skills in a concentrated fashion, so that the result can be graded, and inferences made from the standard of performance that can be expected from the candidate, either at the time of the test or at some future time.

As Carrol (1965, p.364) points out:

The purpose of testing is always to render information to aid in making intelligent decisions about possible courses of action. Sometimes these
decisions affect only the future design or used of the tests themselves, in which case we are dealing with solely experimental uses of tests. Sometimes the decisions have to do with the retention or alteration of courses of training, as when one decides that poor tests results are due to in effective training.

Pit Corder (1973) is of the view that:

Language tests are applied to learners and not to the teaching, teachers or instructional materials. For this reason they do not tell us ‘directly’ about the contribution of the ‘teacher’ or the ‘materials’ to the learning process. They are designed to measure the learners ‘knowledge of’ or ‘competence’ in the language at particular moment in his course and nothing else. The knowledge of one pupil may be compared with the knowledge of others or with that of the same pupil at a different time, or with same standard or norm, as in the case of height, weight, temperature etc. (p.351)

Halliday, et al., (1966, p.215) state that “tests are an attempt to construct an instrument for measuring attainment, or progress, or ability in language skills”. Thus testing, through techniques and skill, provides an estimate of learning, teaching, students’ problems and students’ progress against the percentile.

According to Bailey (1998), “testing is a method to assess a person’s linguistic ability involving a set of techniques, procedures and test items, expecting an active participation and performance of both the examinee and the examiner” (cited in Dikli, 2003, 15). Fulcher (2010) adds “but these decisions are diverse, and need to be made very specific for each intended use of a test” (p.1).

Cronbach (1984) puts this more succinctly:

A test is selected for a particular situation and purpose. What tests are pertinent for a psychological examination of a child entering first grade? That depends on what alternative instructional plans the school is prepared to follow. What test of skill in English usage is suitable for surveying a high school class? Those teachers for whom clarity of expression is important will be discontented with a test requiring only that the student choose between grammatically correct and incorrect expressions. (as cited in Fulcher, 2010, 2)

Messick (1989, p.86–87) argues that testing, when done well, is capable of delivering ‘distributive justice’.
‘Testing’ can be differentiated with another common term ‘examination’. Measurement of students’ competence is referred as ‘test’, whereas achievement of a particular standard after a level of the students is referred as ‘examination’. Both ‘tests’ and ‘examinations’ have the same common function, namely evaluation, therefore, it has become conventional to call them ‘tests’ (Rahman & Gautam, 2012).

To make the meaning of ‘testing’ clearer, it can be compared with another common term ‘assessment’. While ‘testing’ is more formal and standardized; ‘assessment’ is based on a collection of information about what students know and what they are able to do” (Dikli, 2003, 13). In testing, student knows the procedure of scoring, while in assessment the methods differ for collecting information according to different context and time (Law & Eckes, 1995). Testing can be defined as single-occasion, un-dimensional, timed exercises usually in multiple choice or short answer form, while ‘assessment’ is used more widely (Kuliek, et al, 1990).

In testing, however, we are almost always interested in quantifying mental attributes and abilities, sometimes called traits or constructs, which can only be observed indirectly. These mental attributes include characteristics such as aptitude, intelligence, motivation, field dependence/independence, attitude, native language, fluency in speaking, and achievement in reading comprehension (Bachman, 1990, 19).

Carroll (1968) states “a psychological or educational test is a procedure designed to elicit certain behaviours from which one can make inferences about certain characteristics of an individual” (p.46).

By the above definition, one can say that a test is a measurement instrument designed to elicit a specific sample of an individual’s behaviour. What distinguishes a test from other types of measurement is that it is designed to obtain a specific sample of behaviour. While the primary function of ‘assessment’ is to develop a sense of improvement in the learners rather than simply grading and marking their performance. Grading comes later; the main purpose is to identify the problems of learners and subsequently improving their learning and performance. ‘Assessment’, therefore, focuses on the learning of students and results (marks/grades) may be used as a source of information for evaluation (STRIDE, 2003, p. 27).

Over centuries, language testing has adopted various types of tests, strategies and techniques under the influence of changing primacy and focus of different language
theories, approaches and methods. In this section, therefore, the basic principles and some major types of language testing have been briefed:

3.1.1 Principles of Language Testing

Language testing is mainly premised on the following principles:

3.1.1.1 Practicality

A good test should be usable and practical in nature. That is, a test should be such that it can be conveniently conducted in a given situation. Some features of practical test are as follows:

- It should be economical in time and money.
- It should be easy in administrating and scoring the responses.
- It should be easy for the interpretation purpose too.

Thus, it is necessary to check the feasibility of a test in terms of cost in budget limit, scoring system, time limit for both teachers in evaluation and students to complete within the determined time frame and administrative details (Brown, 2003; Hughes, 2003).

Brown (2003, 26) further elaborates:

Practicality refers to logistical, down-to-earth, administrative issues involved in making, giving, and scoring an assessment instrument, these include “costs, the amount of time it takes to construct and administer, ease of scoring, and ease of interpreting/reporting the results”(Mousavi, 2009). A test that fails to meet such criteria is impractical.

3.1.1.2 Reliability

The term Reliability has been defined by Longman Dictionary of Language Teaching & Applied Linguistics as follows:

Reliability measures “the degree to which a test gives consistent results. A test is said to be reliable if it gives same results, when it is given on different occasions or when it is used by different people” (Richards, and Schmidt, 2010, p. 495).

Popham (2009, p.25) defines, “Assessment reliability refers to the consistency with which a test measures whatever it is measuring. It sounds simple enough. In fact, if
Aspects of Formative Assessment

you remember that reliability = consistency, you have mastered most of what you need to know about assessment reliability.”

Brown (2003) explains that “a reliable test is consistent and dependable. If you give the same test to the same student or matched students on two different occasions, the test should yield similar results” (p.27).

He further elaborates its features in the following:

A reliable test

- Is consistent with its conditions across two or more administrations
- Gives clear directions for scoring/evaluation
- Has uniform rubrics for scoring/evaluation
- Lends itself to consistent application of those rubrics by the scorer
- Contains items/tasks that are unambiguous to the test-taker (27).

Reliability employs for the students, the test administration, the test itself, and the teacher. A reliable test has to be reliable to the test and the scorer, as explained below:

**Test Reliability**: Test reliability is achieved by conducting sampling of tasks on varieties of students. Sometimes test reliability is influenced by the change in administration or the lack of motivation among students.

**Scorer Reliability**: It concerns the stability or consistency with which test performances are evaluated. Scorer reliability is frequently influenced in subjective tests or impressionistic evaluations on such test items like composition task, or free response tests, and so forth. To achieve this reliability, an endeavour is regularly made to build up a scoring scale to be trailed by all evaluators.

**3.1.1.3 Validity**

*Longman Dictionary of Language Teaching & Applied Linguistics* (Richards and Schmidt, 2010), defines validity as:

The degree to which a test measures, what it is supposed to measure, or can be used successfully for the purposes for which it is intended. A number of different statistical procedures can be applied to a test to estimate its validity. Such procedures generally seek to determine what the test measures, and how well it does so. (p.622)
Popham (2009, p.30) Claims:

...essentially, validity = inference accuracy. The term valid does not refer to an attribute of tests themselves but, rather, to the accuracy of test-based inferences individuals make based on the test’s results. So, even though we often hear people talking about “a valid test,” this is technically incorrect. It is not a test that is valid or invalid; it is the score-based inference we make about the test results that is valid or invalid.

**Construct Validity**

Recently, the general notion of validity has been replaced by ‘construct validity’ since constructs like reading ability, fluency in speaking, accuracy in writing, written composition, control of grammar and others are generally measured.

*Construct validity* (Richards and Schmidt, 2010, p.124) is:

A type of validity that is based on the extent to which the items in a test reflect the essential aspects of the theory on which the test is based (i.e., construct). For example, the greater the relationship that can be demonstrated between a test of communicative competence in a language and the theory of communicative competence, the greater the construct validity of the test.

Bachman and Palmer (1981) made the early attempts to introduce construct validity to language testing. Cripper and Davies (1988) and Hughes, Porter and Weir (1988) made further significant contributions on these lines.

Nonetheless, it is not sufficient to say that a test needs to have construct validity. Rather it is required to reflect the construct validity in the form of some evidences, like ‘Content validity’, ‘Criterion-related validity’, and ‘face validity’.

**Content Validity**

This dictionary defines *Content validity* as:

A type of validity that is based on the extent to which a test adequately and sufficiently measures the particular skills or behavior it sets out to measure. For example, a test of pronunciation skills in a language would have low content validity if it tested only some of the skills that are required for accurate pronunciation, such as a test that tested the ability to pronounce isolated sounds, but not stress, intonation, or the pronunciation of sounds within words.
Content validity is of particular importance in criterion-referenced tests, where the test content must represent the content of what has been taught in a course. (Richards and Schmidt, 2010, p.126)

It involves analysis of content and making judgments on subject matter to determine the adequacy and representativeness of the test content, i.e., the area to which test results are intended to test (Weigle, 2002; Greenberg, 1992). It has been often observed that the test designers ignore the ‘content validity’ and give preference to simpler and easier content over the important content.

**Criterion-related validity:**

*Criterion-related validity* is:

A type of validity that is based on the extent to which a new test is compared or correlated with an established external criterion measure. For example a new test of L2 vocabulary can be validated by correlating the test score of the new test with that of some other criterion measure representing an identified construct (i.e. L2 vocabulary Knowledge). (Richards and Schmidt, 2010, p.144)

There are two types of criterion validity like concurrent validity and predictive validity:

**Concurrent validity**

*Concurrent validity* is:

A type of validity based on the extent to which a test correlates with some other test that is aimed at measuring the same skill, or with some other comparable measure of the skill being tested. For example, to determine the concurrent validity of a new L2 listening comprehension test, one could calculate the correlation between scores of a group of test takers on this test with their scores on an existing valid and reliable test of L2 listening comprehension at about the same time. (Richards and Schmidt, 2010, p. 114)

**Predictive Validity**

based on the degree to which a test accurately predicts future performance” (451).

Consequential Validity

Consequential validity is “a type of validity that is based on the extent to which the uses and interpretation of a test that may have an impact on society will result in fair and positive social consequences for all stakeholders including test takers” (Richards and Schmidt, 2010, p.119).

Face Validity

Face validity

The degree to which a test appears to measure the knowledge or abilities it claims to measure, based on the subjective judgment of an observer. For example, if a test of reading comprehension contains many dialect words that might be unknown to the test takers, the test may be said to lack face validity. (Richards and Schmidt, 2010, p. 215)

3.1.1.4 Authenticity

Authenticity is “the extent to which test tasks correspond to language use in a non-test (i.e. target language use) situation”. (Richards and Schmidt, 2010, p.43).

It is defined as “the degree of correspondence of the characteristics of a given language test task to the features of a target language use (TLU) task” (Bachman and Palmer, 1996, p. 23).

3.1.1.5 Washback

Longman Dictionary of Language Teaching Applied Linguistics, (Richards and Schmidt, 2010), defines it as “the effect of a test on individual test takers, other stakeholders (e.g. teachers, parents, school administrators, or test developers), educational systems, or society”(p.272)

This refers to effect of tests on individual test taker and teacher, educational institutions to society. Bachman and Palmer (1996) argue:

Tests are never administered in vacuum. Tests reflect the societal concepts, values and the test scores have consequences for individuals and group that
must be considered carefully in making decisions regarding the administration and of tests. (as cited in Weigle, 2002, p. 54)

This effect of test on educational modules and guideline is eluded as wash back impact. Wash back can have positive and negative impact on tests. Positive wash back impact can be characterized as any impact of the testing that urges educator to utilize best current practices in the instructional method documented.

Wash-back is a complex process and results are not always straight forward. The design of test cannot guarantee the positive wash back effect. But beneficial positive wash back can occur if a test taker, teacher and administration understand the purpose, goals of the test that should give the informative and reliable results. The test should involve the use of with clear assessment procedure (Bailey, 1996 cited in Weigle, 2002).

3.1.2 Types of Language Tests

In order to meet various requirements and purposes, series of tests were designed and developed over a period of time, some of which are as follows:

3.1.2.1 Language Aptitude Test

Language aptitude test “measures a person’s aptitude for second language or foreign language learning and that can be used to identify those learners who are most likely to succeed” (Richards and Schmidt, 2010, p. 313).

Popham(2009, p.43) argues:

…aptitude tests were thought to measure a student’s potential, according to aptitudes usually believed to be genetically inherited, achievement tests were presumed to measure what a student had taught—typically the sorts of knowledge and skills that have been taught in school. In recent years, aptitude tests have been increasingly criticized on the grounds that they are not only too genetically dependent but also insufficiently predictive… Even so, teachers should be familiar with the distinction between tests designed to measure aptitude and tests designed to measure achievement.

In other words, aptitude tests are used to predict applicants’ success in achieving certain objectives in the future. In order to take an aptitude test, the examinee does not need to have prior knowledge of the subject being tested. In language education,
through aptitude tests, one can determine which language the examinee is likely to learn more easily and successfully than other language (Farhadi, Ja’farpur and Birjandi, 1994).

According to Brown (2003, p. 43), an Aptitude test “is designed to measure capacity or general ability to learn a foreign language and ultimate success in that undertaking. Language aptitude tests are ostensibly designed to apply to classroom learning of any language.”

### 3.1.2.2 Proficiency Test

A proficiency test is one that measures “how much of a language someone has learned… not linked to a particular course of instruction, but measures the learner’s general level of language mastery” (Richards and Schmidt, 2010, p. 464).

The goal of proficiency test is to weigh the overall, global talent in a language in order that the learners can be assessed if they have had enough capability to deal with the language in a given situation, like completing a certain course in the US or the UK. IELTS and TOEFL are such proficiency tests.

### 3.1.2.3 Placement Test

*Longman Dictionary of Language Teaching and Applied Linguistics*, (Richards and Schmidt, 2010), defines *placement test* as “a test that is designed to place test takers an appropriate level in a program or course. The term “placement test” does not refer to what a test contains or how it is constructed, but to the purpose for which it is used” (440).

The main purpose of placement test is to place the learner at a correct level in course of study. This is conducted at the beginning of the course study (brown, 2003; Hughes, 2003).

### 3.1.2.4 Diagnostic Test

*Diagnostic test* is:

A test that is designed to provide information about L2 learners’ strengths and weaknesses…Diagnostic tests may be used to find out how much L2 learners know before beginning a language course to better provide an efficient and effective course of instruction. (Richards and Schmidt, 2010, p. 168)
The objective of Diagnostic is to examine the learners’ strength and weakness in a specific area of language learning. Meanwhile it guides teachers to recognize future need of learners in that specific area of course study. On the other hand it is supportive for the improvement of students’ knowledge during the educational program. In addition it provides helpful insights for teacher to adapt their teaching methodology in respect to the results of the taken test. (Brown, 2003; Hughes, 2003).

3.1.2.5 Achievement Test

*Achievement test* is:

A test designed to measure how much of a language learners have successfully with specific reference to a particular course, textbook, or program of instruction, thus a type of criterion-referenced test. An achievement test is typically given at the end of a course, whereas when administered periodically throughout a course of instruction to measure language learning up to that point, it is alternatively called a Progress test. Its results are often used to make advancement or graduation decisions regarding learners or judge the effectiveness of a program, which may lead to curricular changes. (Richards and Schmidt, 2010, p. 6)

This test is related to the goals of the course content or curriculum. The Progress test can also be understood in terms of ‘Achievement’ test. This test is closely related to the curriculum and they establish exactly what a learner has learned in a given teaching context. That is why; most teachers are involved in either framing or evaluating this test. The achievement test can be categorized into two types: ‘Final Achievement Test’ and ‘Progress Achievement Test’. These tests too aim at better performance and positive washback (Brown, 2003; Hughes, 2003).

3.2 Evaluation

Each term has its own practice and purpose in the area of language learning and teaching. ‘Evaluation’ is the procedure concerned with decision-making policy to determine the quality of the present performance of client often compared to set standards to make judgment (Baehr, 2010, p.8). In other words, ‘evaluation’ examines whether a standard was met to the set criteria to observe the outcome during or after the learner’s performance. Then evaluation results have no direct bearing on students’ assessment.
Aspects of Formative Assessment

*Longman Dictionary of Applied Linguistics (Richards, Platt and Weber (eds.))* defines ‘Evaluation’ as:

…in general, the systematic gathering of information for the purposes of decision making. Evaluation uses quantitative methods (e.g. tests), qualitative methods (e.g. observations, ratings) and value judgments. In Language Planning, evaluation frequently involves gathering information on patterns of language use, language ability, and attitudes towards language. In language teaching programmes, evaluation is related to decisions to be made about the quality of the programme itself, and decisions about individuals in the programmes. The evaluation of programmes may involve the study of curriculum, objectives, materials, and tests or grading systems… In evaluating both programmes and individuals, tests and other measures are frequently used.” (Richards, Platt and Weber (eds.), p.98).

Kizlik (2012) explained in his study:

*Evaluation* is perhaps the most complex and least understood of the terms. Inherent in the idea of evaluation is ‘value’. When we evaluate, what we are doing is engaging in some process that is designed to provide information that will help us make a judgment about a given situation. Generally, any evaluation process requires information about the situation in question. A situation is an umbrella term that takes into account such ideas as objectives, goals, standards, procedures, and so on. When we evaluate, we say that the process will yield information regarding the worthiness, appropriateness, goodness, validity, legality, etc., of something for which a reliable measurement or assessment has been made. … We evaluate every day. Teachers, in particular, are constantly evaluating students, and such evaluations are usually done in the context of comparisons between what was intended (learning, progress, behaviour) and what was obtained. (p.2)

Tharp and Gallimore (1988) argue that ‘evaluation’ is not synonymous with assessment, because evaluation is the procedure of assessing values to the learning outcomes during and at the end of a course. ‘Evaluation’ is “the process of judging the quality or value of a performance or a course of action” (Airasian, Engemann, & Gallagher, 2007, p. 2).
### 3.3 Assessment

Assessment is a recent phenomenon, which is used broadly and more commonly in the area of language teaching these days. It refers to activities which include purposeful techniques of interaction so as to obtain the understanding of what learners have achieved. The essential point of ‘assessment’ is to facilitate feedback to the learners’ needs, teachers’ need, institution’s needs, societal needs as well.

*Longman Dictionary of Applied Linguistics, (Richards, Platt and Weber (eds.)),* explains *Assessment* as “the measurement of the ability of a person or the quality or success of a teaching course, etc.” It further explains that “assessment may be by test, interview, questionnaire, observation, etc” (p.35).

‘Assessment’ also refers to “the process of collecting, synthesizing, and interpreting information to aid in decision making” (Airasian, Engemann, & Gallagher, 2007, p.2). According to Angelo (1995),

> Assessment is an ongoing process aimed at understanding and improving student learning. It involves making our expectations explicit and public; setting appropriate criteria and standards for learning quality; systematically gathering, analyzing, and interpreting evidence to determine how well performance matches those expectations and standards, and using the resulting information to document, explain and improve performance. (Pp.7-9)

Dylan Wiliam defines *Assessment* in his personal communication with Popham,

> Assessment is the key process in instruction. It is only through assessment that we can find out whether what has been taught has been learned. Assessment is, therefore, the bridge between teaching and learning. (Personal communication, November 7, 2006, as cited in Popham, 2009, p. 101)

Heritage, M (2007, 140) explains, “To many of today’s teachers, assessment is synonymous with high–stakes standardized tests. But there is an entirely different kind of assessment that can actually transform teaching and learning.”

Broadly speaking, there are two types of Assessment: Formative and Summative.

#### 3.3.1 Formative Assessment

Formative Assessment has been defined variously. Some of them are as follows:
Longman Dictionary of Applied Linguistics” (Richards and Schmidt, 2010), defines Formative Evaluation as, “the process of providing information to curriculum developers during the development of a curriculum or program, in order to improve it. Formative evaluation is also used in syllabus design and development of language teaching program and materials” (p.227).

Bloom et al., (1971) claim that

Formative assessment is for us the use of systematic evaluation in the process of curriculum construction, teaching, and learning for the purpose of improving any of these three processes. Since formative assessment takes place during the formation stage, every effort should be made to use it to improve the process. This means that in formative evaluation one must strive to develop the kinds of evidence that will be most useful in the process, seek the most useful method of reporting the evidence, and search for ways of reducing the negative effect associated with evaluation - perhaps by reducing the judgmental aspects of evaluation or, at the least, by having the users of the formative evaluation...make the judgment, (Pp. 17-18)

Cizek (2010) added a meaningful definition as follows:

Formative assessment refers to the collaborative processes engaged in by educators and students for the purpose of understanding the students’ learning, and conceptual organization, identification of strengths, diagnosis of weaknesses, areas for improvement, and as a source of information that teachers can use in instructional planning and students can use in deepening their understandings and improving their achievement. (p. 6)

According to Palomba and Banta (1999), Formative Assessments are:

Observations which allow one to determine the degree to which students know or are able to do a given learning task, and which identifies the part of the task that the student does not know or is unable to do. Outcomes suggest future steps for teaching and learning...Assessment conducted during a performance/course/program with the purpose of providing feedback that can be used to modify, shape, and improve a performance/course/program. (p.7)
Popham (2008, p.5) clarifies Formative Assessment for Students and Teachers (FAST) State Collaborative on Assessment and Student Standards (SCASS) Group definition that

“Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students’ achievement of intended instructional outcomes”…by describing the key features of the FAST SCASS definition:

• Formative assessment is a process, not any particular test.
• It is used not just by teachers but by both teachers and students.
• Formative assessment takes place during instruction.
• It provides assessment-based feedback to teachers and students.
• The function of this feedback is to help teachers and students make adjustments that will improve students’ achievement of intended curricular aims.

Popham (2008, p.112) Further defines, “Formative assessment is a planned process in which assessment-elicited evidence of students’ status is used by teachers to adjust their ongoing instructional procedures or by students to adjust their current learning tactics.”

Black and Wiliam (1998b) contain a ‘what’ element and a ‘when’ element’ within their notion of FA:

‘What?’: all those activities undertaken by teachers, and by their students in assessing themselves, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged.
‘When?’ ‘Such assessment becomes 'formative assessment' when the evidence is used to adapt the teaching work to meet the needs. (p. 2)

Black et al., (2002) further elaborate it when he exclaims:

An assessment activity can help learning if it provides information to be used as feedback, by teachers, and by their [students] in assessing themselves and each other, to modify the teaching and learning activities in which they are engaged. Such assessment becomes ‘formative assessment’ when the evidence
Aspects of Formative Assessment

is actually used to adapt the teaching work to meet learning needs. (as cited in Black, 2007, p.1)

Formative assessment is defined as “information communicated to the learner that is intended to modify his or her thinking or behaviour for the purpose of improving learning” (Shute, 2008, p.154)

Shepard (2008) provides a more comprehensive definition when argue that “Formative assessment is defined as assessment carried out during the instructional process for the purpose of improving teaching or learning. . . . What makes formative assessment formative is that it is immediately used to make adjustments so as to form new learning” (p.281).

In other words it can be stated that “An assessment is formative to the extent that information from the assessment is fed back within the system and actually used to improve the performance of the system in some way” (Wiliam & Leahy, 2007, p. 31).

Sadler (1998, p. 77) rightly claims, “[Formative assessment] refers to assessment that is specifically intended to provide feedback on performance to improve and accelerate learning.”

“Formative assessment, therefore, is essentially feedback both to the teachers and to the learners about present understanding and skill development in order to determine the way forward” (Ramaprasad, 1983, as cited in Harlen & James, 1997, p. 369).

Surgenor, p. (2010) further elaborates in the following manner:

Formative assessment does not form part of the student’s final grade or mark. It is used to provide constructive feedback to improve learning and understanding. The product of formative assessment may never be quantifiably recorded on a grade sheet. The rationale is that students learn effectively by making and learning from mistakes which is difficult to do if their academic performance/final grade may be adversely affected. Since there are no marks at risk students can be more experimental, challenging preconceived ideas and developing more desirable higher cognitive skills. Brown & Knight (1994) suggest that ungraded assessment, where no mark is attached, may become the norm, allowing students the freedom to experiment and be more adventurous in their study and exploration of their subject. Despite the fact that no marks are associated with formative assessment,
students do still engage in it if its role in helping students improve their grades is clearly explained. Students will be motivated if they clearly see the point of their work; how it relates to the course, the module, and their career goals; if it is inherently rewarding or interesting; or if they can see their skills and expertise advancing. (p.2)

McManus (2008, p.3) defines, “Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students’ achievement of intended instructional outcomes”.

3.3.2 Summative Assessment

Formative Assessment can be differentiated with Summative Assessment in the following definitions and elaborations by various experts:

*Longman Dictionary of Applied Linguistics* (Richards and Schmidt, 2010) defines “Summative evaluation is the process of providing information to decision makers, after the program is completed, about whether or not the program was effective and successful” (p.227).

Bloom et al., (1971) say:

*We have chosen the term “summative evaluation” to indicate the type of evaluation used at the end of a term, course, or program for purposes of grading, certification, evaluation of progress, or research on the effectiveness of a curriculum, course of study, or educational plan. [...] Perhaps the essential characteristic of summative evaluation is that a judgment is made about the student, teacher, or curriculum with regards to the effectiveness of learning or instruction, after the learning or instruction has taken place, (p. 118)*

McMillan (2007, p.17) claims, “Summative assessment is more formal and occurs after instruction is completed”.

Irons (2008) describes, *summative assessment* as “any assessment activity which results in a mark or grade which is subsequently used as a judgment on student performance” (p. 7).

Summative Assessment, “is done at the conclusion of a course or some larger instructional period (e.g., at the end of the program). The purpose is to determine
success or to what extent the program/project/course met its goals” (Assessment Terminology for Gallaudet University, August 2016).

Surgenor, p. (2010),

Summative assessment is the type with which most people are familiar. It is usually conducted in the last few weeks of term to see how well students have learned what they were supposed to have learned.

The results from these assessments are aggregated and used to determine whether a student has fulfilled the specified learning outcomes and may achieve some kind of accreditation. This usually causes a degree of anxiety since the grades received in summative assessments are final and can affect their future prospects.

In summative assessments, therefore, students are less keen to experiment with ideas and concepts (Biggs & Tang, 2007), preferring to ‘play it safe’, giving the answers they believe are expected of them, and banking the marks provided. (p.2)

3.4 Formative versus Summative Assessment

Formative and Summative Assessment can be better understood by the following comparative definitions, charts and diagrams:

Popham (2009) elaborates the two types of assessments as follow:

Summative assessment refers to the use of assessment-based evidence when arriving at decisions about already-completed instructional events such as the quality of a year’s worth of schooling or the effectiveness of a semester-long algebra course. Summative assessment is intended to help us arrive at go/no-go decisions based on the success of a final-version instructional program.

In contrast, formative assessment is a process in which assessment-elicted evidence is used by teachers to adjust their ongoing instructional activities, or by students to adjust the ways they are trying to learn something. In contrast to its summative sibling, formative assessment has a powerful improvement orientation, because it is intended to stimulate ameliorative adjustments in teachers’ still-malleable instructional programs or in students’ current learning-tactics. (p.5)
Following Bloom et al., (1971), the distinction is typically made between formative and summative assessment, the latter being concerned with determining the extent to which a student has achieved curricular objectives. As a number of writers have observed, the distinction between formative and summative assessment is however far from sharp. Some assessments (e.g. in course assignments) are deliberately designed to be simultaneously formative and summative – formative because the student is expected to learn from whatever feedback is provided, and summative because the grade awarded contributes to the overall grade at the end of the study unit. Summative assessments in relation to a curricular component (the student passes or fails a module, for example) can act formatively if the student learns from them. One of the less desirable effects of the unitisation of curricula in UK higher education has been the reduction in the amount of formal formative assessment as the number of end-of-unit summative assessments has increased. Also feedback is sometimes received too late for student choice and may also be insufficient, if only given as a mark or grade, for learning on subsequent modules. (as cited in Mantaz Yorke, 2003, Pp.479-480).

A better understanding of Summative and Formative Assessment can be made from the following comparative chart:

<table>
<thead>
<tr>
<th>Relation to Instruction</th>
<th>Formative</th>
<th>Summative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occurs during instruction</td>
<td>•</td>
<td>• Occurs after instruction</td>
</tr>
<tr>
<td>Frequency</td>
<td>Occurs on a ongoing basis (daily)</td>
<td>• Occurs at a particular point in time to determine what students know</td>
</tr>
<tr>
<td>Relation to Grading</td>
<td>Not graded – information is used as feedback to students and teachers, mastery is not expected when students are first introduced to a concept</td>
<td>• Graded</td>
</tr>
<tr>
<td>Students Role</td>
<td>Active engagement – self assessment</td>
<td>• Passive engagement in design and monitoring</td>
</tr>
<tr>
<td>Requirements for Use</td>
<td>Clearly defined learning targets that students</td>
<td>• Well designed assessment blue print that outlines</td>
</tr>
</tbody>
</table>
### Aspects of Formative Assessment

<table>
<thead>
<tr>
<th>Understand</th>
<th>the learning targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Clearly defined criteria for success that students understand</td>
<td>• Well designed test items using best practices</td>
</tr>
<tr>
<td>• Use of descriptive versus evaluative feedback</td>
<td></td>
</tr>
</tbody>
</table>

#### Examples

- Better thought of as a process rather than a thing. Examples include observations, interviews, evidence from work samples, paper and pencil tasks
- State assessments, interim assessments, end of unit assessments, common assessments

#### Purpose

- Designed to provide information needed to adjust teaching and learning while they are still occurring
- Designed to provide information about the amount of learning that has occurred at a particular point
- Since summative assessments occur after instruction every few weeks, months, or once a year, they help evaluate the effectiveness of programs, school improvement goals, alignment of curriculum, or student placement in specific

(Author. (no. date), [doe] August 2016)

### 3.5 Origin and Theoretical Premise of Formative Assessment

According to William (2010) there is no special theory behind formative assessment. However, Vygotsky’s (1978) notion of the Zone of Proximal Development (ZPD) from constructivism viewpoint, there is a focal attribute to the formative assessment that classroom culture shifted from teacher-centered to learner-centered.

Though a relatively new term, “Formative Assessment” has been used throughout the history as an effective method, and Socrates is considered as its first practitioner. He extracted responses from students by peppering them with question, and used the
answers to measure their learning. The process used by Socrates is considered by many as the initial feature of formative assessment.

As an idea which became prominent in 1990s formative assessment has evolved over time, from a focus on Formative Evaluation to Formative Assessment and Assessment for Learning. Cronbach (1963) refers to the idea of ‘evaluation’ as a medium to improve curricular program and, Scriven (1967) based on the works of Cronbach, proposes the term “Formative” as a way of clarifying the roles of evaluation. He uses the term to differentiate between the goals of collecting information and how that information is used then.

While Scriven was the first one who used the term “Formative Assessment”, Benjamin Bloom was the first to apply the ideas of the formative versus summative to put forward the concept of mastery learning in educational assessment. A concept which gained recognition in 1970s, in which students are not allowed to move to the next level until they have met the objectives set for the current level.

Formative assessment was widely explored in the next decades. Bloom (1971) expanded his theoretical work by identifying two major features of formative assessment including feedback for students and conditions for all important components of learning. He enlarged his taxonomy of cognitive objectives, Bloom’s taxonomy. This referred to different level of thinking skills from Lower Order of Thinking Skill (LOTS) to Higher Order Thinking Skill (HOTS). In 1988, Terry Crooks investigated the impacts of using assessment practices on students learning and their motivation. Crook (1988) also pointed out that classroom assessment “appear to be one of the most potent forces influencing the evaluation” (p.476). By the same time, Sadler (1989) suggested a theory that placed a great emphasis on the role of student in the process of assessment. He argued that assessment can be most effective, when students encouraged to self monitor the quality of their own work.

The idea of formative assessment was embraced specifically in 1998, when Paul Black and Dylan Wiliam published their findings as “Inside the black box”, after completing a meta-analysis of more than 250 research studies on the topic. The idea of “Inside the black box” led to the widespread application of formative assessment throughout the world.
For further explanation of the goals and roles of summative and formative assessment in 1990s, Assessment Reform Group coined the term “Assessment for Learning”. Black and Wiliams (2009) offered a definition of formative assessment that was built over ten years of work with Assessment Reform Group and Wyne Harlen and Mary James, the members of Assessment Reform Group accentuated the necessity of making a distinction between formative and summative assessment.

And during 1980 and 1990s educational research continued expanding the previously offered ideas and theories and replaced the term “formative evaluation” with the term “Formative Assessment” and then to “Assessment for learning”.

A comprehensive discussion of formative assessment evolution can also be found in the studies by Popham (2008), Greenstein (2010), Trumbull & Lash (2013) and Grant & Gareis (2014/2015).

3.6 Aspects / Features of Formative Assessment

Now that we have referred to various definitions and development of Formative Assessment, we need to identify its basic features and the mechanism.

The FA cycle (Figure 3.1) explains how this type of assessment works. It shows that there are certain goals, which are intended to be achieved through instruction as it generally happens in a lecture-based or teacher dominated teaching methodology. But here in the context of FA, the instruction and its results are measured, based on which feedback is provided and corrective measures are taken in teaching and learning, leading to the better achievement of goals.

![Figure 3.1: Formative Assessment Cycle adapted from (Richards, 2015)](image_url)
Aspects of Formative Assessment

In (Figure 3.2) below, the FA cycle has been further elaborated which shows that goals and objectives are intended to be achieved through ‘Targeted instructions’ and informed teaching, the results of which are put to ‘Data analysis’, where the findings of data are responded properly and carefully for an effective achievement of goals/objectives.

![Formative Assessment Cycle Diagram](image)

*Figure 3.2: Formative Assessment adapted from (Greenstein, 2016)*

The features of FA can be better understood in the following table:

<table>
<thead>
<tr>
<th>DOs</th>
<th>Don’ts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highlight the needs of each student</td>
<td>View all students as being, or needing to be, at the same place in their learning</td>
</tr>
<tr>
<td>Provide immediately useful feedback to students and teachers</td>
<td>Provide feedback weeks or months after the assessment</td>
</tr>
<tr>
<td>Occur as a planned and intentional part of the learning in a classroom</td>
<td>Always occur at the same time for each student</td>
</tr>
<tr>
<td>Focus on progress or growth</td>
<td>Focus solely on a number, score, or level</td>
</tr>
<tr>
<td>Support goal setting within the classroom curriculum</td>
<td>Occur outside of authentic learning experiences</td>
</tr>
<tr>
<td>Answer questions the teacher has about students’ learning</td>
<td>Have parameters that limit teacher involvement</td>
</tr>
<tr>
<td>Reflect the goals and intentions of the teachers and the students</td>
<td>Look like mini-versions of pre-determined summative assessments</td>
</tr>
<tr>
<td>Rely on teacher expertise and</td>
<td>Rely on outsiders to score and analyze</td>
</tr>
</tbody>
</table>
Aspects of Formative Assessment

<table>
<thead>
<tr>
<th>interpretation</th>
<th>results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occur in the context of classroom life</td>
<td>Interrupt or intrude upon classroom life</td>
</tr>
<tr>
<td>Focus on responsibility and care</td>
<td>Focus on accountability</td>
</tr>
<tr>
<td>Inform immediate next steps</td>
<td>Focus on external mandates</td>
</tr>
<tr>
<td>Allow teachers and students to better understand the learning process in</td>
<td>Exclude teachers and students from assessing through the</td>
</tr>
<tr>
<td>general and the learning process for these students in particular</td>
<td>whole learning process</td>
</tr>
<tr>
<td>Encourage students to assume greater responsibility for monitoring and</td>
<td>Exclude students from the assessment process</td>
</tr>
<tr>
<td>supporting their own learning.</td>
<td></td>
</tr>
<tr>
<td>Consider multiple kinds of information, based in a variety of tools or</td>
<td>Focus on a single piece of information</td>
</tr>
<tr>
<td>strategies</td>
<td></td>
</tr>
</tbody>
</table>

(National Council of Teachers of English (NCTE).2013, p. 6)

In the light of the above, the aspects and mechanism of Formative Assessment can be represented in the following diagram:

![Figure 3.3: Aspects of Formative Assessment Adapted from (Author.(n.d). [aisd], August 2016)](image)

This diagram shows how the teacher, the peer and the learner contribute in leading effective learning. In formative assessment, the learning objectives stand clarified, shared and understood by all stakeholders, which are targeted and efforts are made through various tasks and activities to achieve with the help of feedback. Here students function as resource for one another and they are considered to be owners of their own learning.
Formative Assessment is implemented systematically in following steps, as explained in the following diagram:

Figure 3.4: Stages of Formative Evaluation adapted from (Agency for Healthcare Research and Quality [AHRQ], March 2013)

The above diagrams, aspects, elaborations and explanations show that Formative Assessment is a highly systematic aspect of language learning and teaching, which promises a better and effective learning and teaching process.

3.7 Instruments / Techniques of Formative Assessment

Formative Assessment adopts various instruments / techniques for effective learning and teaching, some of which are as follows:

3.7.1 Questioning

Questioning is one of the most common and frequent activities of human discourse, including in the classroom situation. This is generally used by teachers to comprehend learners’ level of learning of a certain task or to provide learners an opportunity to ask/raise questions for clarifications and deeper thinking. This type of practice aims to promote learning by engaging students in question/answer dialogue. It helps learners in participating, in developing their cognitive abilities and thinking skills, and in boosting learners’ interest to answer questions, besides other benefits. Posing questions becomes more significant when its aim is to explore issues that are critical
to the promotion of learners’ knowledge. However, this needs more endeavors for framing questions are worth for asking (Black et al., 2003).

3.7.2 Discussion

Discussion engages students in a group conversation to share their opinions and attitudes regarding a topic. There is an interaction among either students’ group or the whole class. During discussion, students can express their views and expand their knowledge. This technique opens a space for students to become learning resources for their peers so as to improve their understandings and activate their creative and critical thinking too. This can be done as individuals, in pairs or in groups. (Black and Wiliam, 1998).

Discussion based activities trains learners to be participative, to communicate logically, to contribute their understanding and world view / perception, and involves learners in solving problems, exploring issues and thinking deeper. It also helps teachers in evaluating whether learners have acquired knowledge of essential concepts.

3.7.3 Self Assessment

Self assessment is a way of assessing ones’ own learning in order to be aware of his/her strengths and weaknesses. This technique engages students in a meta-cognitive manner to lead them become more effective learners, because they think and compare themselves with their peers. This helps them develop their skills and articulate their understandings.

This activity makes learners motivated, self aware, responsible and self reliant. They learn to identify, diagnose and prescribe their own problems of learning.

3.7.4 Peer Assessment

Peer assessment is a way of assessment by the peer group in the classroom. Peer assessment requires teachers’ support to make a comfortable and trustful environment for the purpose of gaining constructive feedback from peers. This is a giving and receiving feedback practices.

This type of assessment is very useful for large classes and also for time management. Peer assessment provides an opportunity for students to reinforce their learning experience during estimating other students’ efforts in a same task. Therefore they become autonomous.
3.7.5 Practice Presentation

Practice presentation is an absolute student centered activity that prepares students to afford in real life situation. This tool of communication improves students’ communicative competence. Four skills (listening, speaking, reading and writing) are involved in practice presentation. This is a performance-based activity that the audience and peer groups contribute to assess the practitioner’s behavior before final presentation. This procedure is helpful for both participants and practitioners in the classroom. Hence, Practice presentation requires efforts in planning and allotting class time.

Practice presentation improves learners’ research skills, life skills, communication skills, besides developing creative, cognitive and critical abilities.

3.7.6 Visual Representation

Visual representation is a means to clarify the meaning of subject matter for students to understand better. For the purpose of FA, employing visual representation in various forms such as maps, graphs and pictures enhance students’ ability to construct their knowledge. Good visual representation will make learning and assessing process smoother.

It is actually an aid for conveying information in an effective and appropriate manner in order to elaborate concepts and ideas.

3.7.7 Think-Pair-Share

This classroom assessment technique can be used in any content area of teaching a lesson/topic. This assessment technique entails three steps. In first step, teachers ask a question and wait. In next step, Students think silently and then get into pairs and discuss the question to find the answers. In last step, each pair shares their understandings to whole class. In other words, teachers can employ this formative assessment practice when they walk around classroom and observe students’ behavior to gain information about students.

Think-Pair-Share is a technique that provides collaborative learning in classroom setting. It motivates students to participate in a discussion activity with their pairs, which consequently promotes critical thinking and provides all learners an opportunity to participate.
3.7.8 Oral Examination

This technique usually employs within graduate, undergraduate or postgraduate course program that students are interrogated/interviewed by more than one examiner to ensure students’ writing in final examination. Oral Examination can be in any type of assessment that is based on spoken ability of students including viva-voce, oral presentations, etc. Its purpose is to introduce oral assessment into the course in order to make students to express their knowledge orally.

3.7.9 Assignment

Assignment is another one of the most common and useful instructional technique that can be used in any subject area of teaching. Assignment includes writing skills, self-learning and searching for resources.

Assignment improves student’s authorship ability. This activity leads students to be more autonomous in their learning, to apply various study skills, planning, and also make use of the communicative skills effectively.

3.7.10 Project work

According to Richards & Schmidt (2010):

This activity focuses on the completion of a task, and usually requires an extended amount of independent work both for an individual student and a group of students. Much of this work takes place outside the classroom environments. Project works often includes three steps:

1. Classroom planning. The students and teacher talk about the content and scope of work, and their wants.

2. Implementing the project. The students move out of the classroom walls for the completion of planned tasks (e.g. conducting interviews, gathering information).

3. Reviewing and monitoring. This involves discussions and feedback sessions by the teacher and participants, both during and after the project. (Pp. 467-8)

Project work promotes cooperative learning, enhances learner-centered teaching and provides the learners an opportunity to deal with authentic and real life situations and communications.
3.7.11 Portfolio

Pierce and O’Malley (1992) define portfolio as, “the use of records of a student’s work over time and in a variety of modes to show the depth, breadth, and development of the student’s abilities” (p.2).

Additionally, Paulson etal., (1991) define portfolio as:

A portfolio is a purposeful collection of student work that indicates the students’ attempts, progress, and achievements in one or more particular areas of curriculum. The collection needs to consist of the following:

- Student participation in selecting contents.
- Criteria for selection.
- Criteria for judging merits.
- Evidence of a students’ self-reflection.

It should represent a collection of students’ best work efforts, student-selected samples of work experiences related to outcomes being assessed, and documents according growth and development toward mastering identified outcomes. (Pp.60-63)

Portfolio advocates self learning, expands the view of learning and assessing in a continuous process simultaneously, and indicates students’ growth during teaching-learning process.

3.8 Advantages and Disadvantages of Formative Assessment

One may list any number of advantages and disadvantages of Formative Assessment. From practical point of view, the following may be listed:

Advantages: Formative Assessment

1. Provides feedback on teaching, learning and overall language program.
2. Focuses on higher level of knowledge and abilities rather than on measuring basic skills.
3. Makes the classroom interactive through various types of tasks and activities.
4. Motivates students to participate and interact.
5. Trains students in the effective use of study skills.
6. Makes students independent learners.
7. Helps in managing large classrooms through pair and group works.
8. Implements various techniques of assessment to extract information on learning.

9. Provides continuous and systematic process for employing assessment outcomes to develop teaching, learning and curriculum.

10. Makes students responsible for their learning.

11. Focus on teachers’ assessment literacy and developmental training to improve instruction.

12. Provides teachers an opportunity to be innovative.

13. Provides autonomy to teachers to adopt appropriate strategies and techniques for effective learning and teaching.

Disadvantages: Formative Assessment

1. Is time consuming, if not systematically planned and if the context and students are not supportive.

2. Expects learners to be very active, participative and highly motivated.

3. Expects highly motivated and trained teachers.

4. Needs more efforts and planning activities on the part of the teachers.

5. Is not reliable because it is subjective in nature.

6. Expects teachers to possess professional skills in selecting appropriate techniques and strategies.

7. May be difficult to motivate learners to complete the task.

8. May not be practically possible to implement due to lack of proper infrastructure and motivated students.

9. May be difficult for heterogeneous group of students.

10. May face difficulty in scoring, which may consequently lead to low reliability.

11. May face problems in semi-urban or rural regions in the implementation of esp. such FA techniques like Project work, Visual Representation, and Portfolio making.

3.9 Summing Up

The present chapter has attempted to define Formative Assessment and its related terms like ‘testing’, ‘evaluation’, ‘summative assessment’ and others. In addition, it also provides a discussion on the origin, basic principles, features and aspects of Formative Assessment. It has been ensured that this chapter presents a clear
understanding of Formative Assessment and also establishes a background for the data based case study at AMU, which has been dealt with in the next chapter.

While meeting the first two objectives / research questions indicated in Chapter 1, this chapter

i. understands and discusses definitions, features and aspects of Formative Assessment; / What is Formative Assessment? What are its characteristic features? How is it different from the Traditional Pen and Paper Method of assessment?

ii. discusses theoretical premise of Formative Assessment; / To what extent theoretical constructs relating to learning, teaching and assessment can be made effective in the Undergraduate ESL classroom at AMU?
REFERENCES


Aspects of Formative Assessment


Aspects of Formative Assessment


Popham, W.J. (2009). Instruction that measures up: successful teaching in the age of
accountability. Virginia, USA: ASCD. Alexandria.


Aspects of Formative Assessment


