Assessment and Evaluation in ELT

Tomás de Aquino Caluyua Yambi

ISCED, Angola

INTRODUCTION

Unleashing the potential of continuous improvement in teaching/learning in English Language Teaching—ELT requires an appreciation of the difference in spirit between assessment and evaluation. Assessment is frequently confused and confounded with evaluation. The purpose of an evaluation is to judge the quality of a performance or work product against a standard. The fundamental nature of assessment is that a mentor values helping a mentee and is willing to expend the effort to provide quality feedback that will enhance the mentee's future performance. While both processes involve collecting data about a performance or work product, what is done with these data in each process is substantially different and invokes a very different mindset. This paper first looks at what assessment is and the various aspects involving. Then attention will be turn to evaluation and its components. Furthermore, it will look at testingas a tool used by both assessment and evaluation, lastly some differences between assessment and evaluation will be presented.

EPISTEMOLOGY OF ASSESSMENT AND EVALUATION

Assessment and Evaluation are two different concepts with a number of differences between them starting from the objectives and focus. Before we go into details about these differences that set assessment and evaluation apart, let us first pay attention to the two words themselves. According to the Webster Dictionary (2017), assessment means appraisal. Then, according to the same dictionary, evaluation is estimation or determining the value of something. So, these processes are used in the field of ELT very often to test the quality of teaching and learning processes. That is done to let the educational institutes find out what more can be done to improve the education offered by those educational institutes.

WHAT IS ASSESSMENT

As stated above, and according to Brown, (1990) assessment refers to a related series of measures used to determine a complex attribute of an individual or group of individuals. This involves gathering and interpreting information about student level of attainment of learning goals.

Assessments also are used to identify individual student weaknesses and strengths so that educators can provide specialized academic support educational programming, or social services. In addition, assessments are developed by a wide array of groups and individuals, including teachers, district administrators, universities, private companies, state departments of education, and groups that include a combination of these individuals and institutions.

In classroom assessment, since teachers themselves develop, administer and analyze the questions, they are more likely to apply the results of the assessment to their own teaching. Therefore, it provides feedback on the effectiveness of instruction and gives students a measure of their progress. As Brown (1990) maintains,

two major functions can be pointed out for classroom assessment: One is to show whether or not the learning has been successful, and the other one is to clarify the expectations of the teachers from the students (Brown, 1990).

Assessment is a process that includes four basic components:

1) Measuring improvement over time.

2) Motivating students to study.

3) Evaluating the teaching methods.

4) Ranking the students' capabilities in relation to the whole group evaluation.

Why Assessment is Important

First and foremost, assessment is important because it drives students learning (Brown 1990). Whether we like it or not, most students tend to focus their energies on the best or most expeditious way to pass their 'tests.' Based on this knowledge, we can use our assessment strategies to manipulate the kinds of learning that takes place. For example, assessment strategies that focus predominantly on recall of knowledge will likely promote superficial learning. On the other hand, if we choose assessment strategies that demand critical thinking or creative problem solving, we are likely to realize a higher level of student performance or achievement. In addition, good assessment can help students become more effective self-directed learners (Darling-Hammond 2006). As indicated above, motivating and directing learning is only one purpose of assessment. Well-designed assessment strategies also play a critical role in educational decision-making and are a vital component of ongoing quality improvement processes at the lesson, course and/or curriculum level.

Types and Approaches to Assessment

Numerous terms are used to describe different types to learner assessment. Although somewhat arbitrary, it is useful to these various terms as representing dichotomous poles (McAlpine, 2002).

Formative   <————————————>   Summative

Informal   <————————————>   Formal

Continuous <————————————-> Final

Process   <————————————>   Product

Divergent   <————————————>   Convergent

Formative vs. Summative Assessment

Formative assessment is designed to assist the learning process by providing feedback to the learner, which can be used to identify strengths and weakness and hence improve future performance. Formative assessment is most appropriate where the results are to be used internally by those involved in the learning process (students, teachers, curriculum developers). Summative assessment is used primarily to make decisions for grading or determine readiness for progression. Typically summative assessment occurs at the end of an educational activity and is designed to judge the learner's overall performance. In addition to providing the basis for grade assignment, summative assessment is used to communicate students' abilities to external stakeholders, e.g., administrators and employers (Darling-Hammond, 2006).

Informal vs. Formal Assessment

With informal assessment, the judgments are integrated with other tasks, e.g., lecturer feedback on the answer to a question or preceptor feedback provided while performing a bedside procedure. Informal assessment is most often used to provide formative feedback. As such, it tends to be less threatening and thus less stressful to the student. However, informal feedback is prone to high subjectivity or bias. Formal assessment occurs when students are aware that the task that they are doing is for assessment purposes, e.g., a written examination. Most formal assessments also are summative in nature and thus tend to have greater motivation impact and are associated with increased stress. Given their role in decision-making, formal assessments should be held to higher standards of reliability and validity than informal assessments (McAlpine 2002).

Continuous vs. Final Assessment

Continuous assessment occurs throughout a learning experience (intermittent is probably a more realistic term). Continuous assessment is most appropriate when student and/or instructor knowledge of progress or achievement is needed to determine the subsequent progression or sequence of activities (McAlpine 2002). Continuous assessment provides both students and teachers with the information needed to improve teaching and learning in process. Obviously, continuous assessment involves increased effort for both teacher and student. Final (or terminal) assessment is that which takes place only at the end of a learning activity. It is most appropriate when learning can only be assessed as a complete whole rather than as constituent parts. Typically, final assessment is used for summative decision-making. Obviously, due to its timing, final assessment cannot be used for formative purposes (McAlpine 2002).

Process vs. Product Assessment

Process assessment focuses on the steps or procedures underlying a particular ability or task, i.e., the cognitive steps in performing a mathematical operation or the procedure involved in analyzing a blood sample. Because it provides more detailed information, process assessment is most useful when a student is learning a new skill and for providing formative feedback to assist in improving performance (McAlpine 2002). Product assessment focuses on evaluating the result or outcome of a process. Using the above examples, we would focus on the answer to the math computation or the accuracy of the blood test results. Product assessment is most appropriate for documenting proficiency or competency in a given skill, i.e., for summative purposes. In general, product assessments are easier to create than product assessments, requiring only a specification of the attributes of the final product (McAlpine 2002).

Divergent vs. Convergent Assessment

Divergent assessments are those for which a range of answers or solutions might be considered correct. Examples include essay tests. Divergent assessments tend to be more authentic and most appropriate in evaluating higher cognitive skills. However, these types of assessment are often time consuming to evaluate and the resulting judgments often exhibit poor reliability. A convergent assessment has only one correct response (per item). Objective test items are the best example and demonstrate the value of this approach in assessing knowledge. Obviously, convergent assessments are easier to evaluate or score than divergent assessments. Unfortunately, this "ease of use" often leads to their widespread application of this approach even when contrary to good assessment practices. Specifically, the familiarity and ease with which convergent assessment tools can be applied leads to two common evaluation fallacies: the Fallacy of False Quantification (the tendency to focus on what's easiest to measure) and the Law of the Instrument Fallacy (molding the evaluation problem to fit the tool) (McAlpine 2002).

Approaches to Assessment

In approaches to assessment, two central tendencies emerge which are relevant to language as subject. One places emphasis on the assessment *of* learning where reliable, objective measures are a high priority. The focus here is on making summative judgements which in practice is likely to involve more formal examinations and tests with marks schemes to ensure that the process is sound (McAlpine 2002). An alternative approach is to change the emphasis from assessment *of* learning to assessment *for* learning, implying a more formative approach where there is much more emphasis on feedback to improve performance. The approach here might be through course work and portfolio assessment in which diverse information can be gathered which reflects the true broad nature of the subject (McAlpine 2002).

BETWEEN ASSESSMENT AND EVALUATION

After collecting data from students—assessment, there is then the need for assigning students with numbers or others symbols to a certain characteristic of the objects of interest according to some specified rules in order to reflect quantities of properties. This is called measurement and can be attributed to students' achievement, personality traits or attitudes. Measurement then is the process of determining a quantitative or qualitative attribute of an individual or group of individuals that is of academic relevance (Bachman 1995). A test therefore, will serve as the vehicle or measurement instrument used to observe or elicit an attribute whether in a written test or an observation or an oral question or an assessment intended to measure the respondents' knowledge or other abilities. Then if the test is the vehicle or a measurement tool, then the test score is the indication of quantification—what was observed through the test. The only difference between measurement and test is that while measurement *per se* looks at the holistic perspective, a test on the other hand, is usually designed to obtain a specific sample of behavior whether quantitative or qualitative in nature (Bachman, 1995).

A good test should possess not only validity in its various types (Fulcher & Davidson 2007), and reliability but also objectivity, objective basedness, comprehensiveness, discriminating power, practicability, comparability and also utility (Shohamy 1993; Fulcher & Davidson 2007). Objectivity is when a test is to be said objective if it is free from personal biases in interpreting its scope as well as in scoring the responses. It can be increased by using more objective type test items and the answers are scored according to model answers are provided. Objective basedness is that a test should be based on pre-determined objectives. And a test setter should have definite idea about the objective behind each item (Shohamy 1993). Comprehensiveness is that the test should cover the whole syllabus, due importance should be given all the relevant learning materials, and a test should cover all the anticipated objectives. Validity is the degree to which test measures what it is to measure. Reliability is of a test refers to the degree of consistency which it measures what is intended to measure. A test may be reliable but need not be valid. This is because it may yield consistent scores but these scores need not be representing what is exactly measured what we want to measure (Shohamy 2001). Discriminating power of the test is its power to discriminate between the upper and lower groups who took the test. The test should have different difficulty level of questions. Practicality of the test depends on administrative, scoring, interpretative ease and economy. Comparability is when a test possesses comparability when scores resulting from its use can be interpreted in terms of a common base that has a natural or accepted meaning. Then lastly the utility, a test has utility if it provides the test condition that would facilitate realization of the purpose for which it is mean.

Educators believe that every measurement device should possess certain qualities. Perhaps the two most common technical concepts in measurement are reliability and validity (Weir 2005). Any kind of assessment, whether traditional or "authentic," must be developed in a way that gives the assessor accurate information about the performance of the individual (Weir 2005). At one extreme, we wouldn't have an

individual paint a picture if we wanted to assess writing skills. A test high validity has to be reliable also for the score will be consistent in both cases. A valid test is also a reliable test, but a reliable test may not be a valid one (Shohamy 2001).

WHAT IS EVALUATION

It was said earlier that a test is necessarily a measurement tool, but not all measurement tool is considered a test. On the same token, evaluation does not necessarily entail testing. Tests on the other hand, do not necessarily evaluate (Bachman 1995). For Bachman, tests are often used for several puposes, including for pedagogical purposes, for motivating students to study, for reviewing material previously taught, or for purely descriptive purposes. Only when the results of tests are used as a basis for making a decision that evaluation is involved. So, more specifically, in the field of ELT, evaluation means measuring or observing the process to judge it or to determine it for its value by comparing it to others or some kind of a standard (Weir & Roberts, 1994). The focus of the evaluation is on grades. It is rather a final process that is determined to understand the quality of the process. The quality of the process is mostly determined by grades. That is such an evaluation can come as a paper that is given grades. This type of paper will test the knowledge of each student. So, here with the grades, the officials come try to measure the quality of the programme. Furthermore, Evaluation is comparing a student's achievement with other students or with a set of standards (Howard & Donaghue 2015). It refers to consideration of evidence in the light of value standards and in terms of the particular situations and the goals, which the group or individuals are striving to attain. Evaluation designates more comprehensive concept of measurement than is implied in conventional tests and examination. The emphasis of evaluation is based upon broad personality change and the major objectives in the educational program (Howard & Donaghue 2015).

Evaluation can, and should, however, be used as an ongoing management and learning tool to improve learning, including five basic components according to Kizlik (2010):

1) Articulating the purpose of the educational system.

2) Identifying and collecting relevant information.

3) Having ideas that are valuable and useful to learners in their lives and professions.

4) Analyzing and interpreting information for learners.

5) Classroom management or classroom decision making.

Well-run classes and effective programs are those that can demonstrate the achievement of results. Results are derived from good management. Good management is based on good decision making. Good decision making depends on good information. Good information requires good data and careful analysis of the data. These are all critical elements of evaluation.

Functions of evaluations

Evaluation refers to a periodic process of gathering data and then analyzing or ordering it in such a way that the resulting information can be used to determine how effective your teaching or program is, and the extent to which it is achieving its stated objectives and anticipated results (Howard & Donaghue (2015). Teachers can and should conduct internal evaluations to get information about their programs, to know who passes and who fails so that they can make sound decisions about their practices. Internal evaluation should be conducted on an ongoing basis and applied conscientiously by teachers at every level of an institution in all program areas. In addition, all of the program's participants (managers, staff, and beneficiaries) should be

involved in the evaluation process in appropriate ways. This collaboration helps ensure that the evaluation is fully participatory and builds commitment on the part of all involved to use the results to make critical program improvements (Howard & Donaghue 2015).

Although most evaluations are done internally, conducted by local stakeholders, there is still a need for larger-scale, external evaluations conducted periodically by individuals from outside the program or institution. Most often these external evaluations are required for funding and accreditation purposes or to answer questions about the program's long-term impact by looking at changes in demographic indicators such as graduation rate, changes n economy and other levels. In addition, occasionally a teacher may be observed by an external stakeholder with purpose of assessing programmatic or operating problems that have been identified but that cannot be fully diagnosed or resolved through the findings of internal evaluation (Weir & Roberts, 1994).

Principles of Evaluation

Here are some principles to consider for your own classroom summarised from (Weir & Roberts, 1994; Howard & Donaghue 2015; (Kellaghan & Stufflebean 2003):

- Effective evaluation is a continuous, on-going process. Much more than determining the outcome of learning, it is rather a way of gauging learning over time. Learning and evaluation are never completed; they are always evolving and developing.
- A variety of evaluative tools is necessary to provide the most accurate assessment of students' learning and progress. Dependence on one type of tool to the exclusion of others deprives students of valuable learning opportunities and robs you of measures that help both students and the overall program grow.
- Evaluation must be a collaborative activity between teachers and students. Students must be able to assume an active role in evaluation so they can begin to develop individual responsibilities for development and self-monitoring.
- Evaluation needs to be authentic. It must be based on the natural activities and processes students do both in the classroom and in their everyday lives. For example, relying solely on formalized testing procedures might send a signal to children that learning is simply a search for "right answers."

ASSESSMENT VS. EVALUATION

Depending on the area of study, authority or reference consulted, assessment and evaluation may be treated as synonyms or as distinctly different concepts. In ELT, assessment is widely recognized as an ongoing process aimed at understanding and improving student learning. Assessment is concerned with converting expectations to results. It can be a process by which information is collected through the use of test, interview, questionnaire observation, etc. For example, having your students to write on a given topic your are collecting information, this is what we mean here by assessment (Kizlik 2010; Richards and Schmidt 2002; Weir & Roberts, 1994).

Evaluation on the other hand, is recognized as a more scientific process aimed at determining what can be known about performance capabilities and how these are best measured. Evaluation is concerned with issues of validity, accuracy, reliability, analysis, and reporting. It can therefore be seen as the systematic gathering of information for purposes of decision-making, using both quantitative methods (tests) and qualitative methods (observations, ratings and value judgments) with purpose of judging the gathered information. In other words, when the teachers receive written assignment from students, some kind of correction and/or response and a possible mark will be given. Thus we are in presence of evaluation. However, assessment and evaluation are similar in that they both involve specifying criteria and collecting data/information. In

most academic environments, they are different in purpose, setting criteria, control of the process, and response. For example, an instructor can use the results of a midterm exam for both assessment and evaluation purposes. The results can be used to review with the students course material related to common mistakes on the exam (i.e. to improve student learning as in assessment) or to decide what measurement or grade to give each student (i.e. to judge student achievement in the course as in evaluation) (Howard & Donaghue 2015).

KEY DIFFERENCES BETWEEN ASSESSMENT AND EVALUATION

The significant differences between assessment and evaluation are discussed in the points given below summarized from (Weir & Roberts, 1994; Howard & Donaghue 2015; (Kellaghan & Stufflebean 2003):

1. The process of collecting, reviewing and using data, for the purpose of improvement in the current performance, is called assessment. A process of passing judgment, on the basis of defined criteria and evidence is called evaluation.
2. Assessment is diagnostic in nature as it tends to identify areas of improvement. On the other hand, evaluation is judgemental, because it aims at providing an overall grade.
3. The assessment provides feedback on performance and ways to enhance performance in future. As against this, evaluation ascertains whether the standards are met or not.
4. The purpose of assessment is formative, i.e. to increase quality whereas evaluation is all about judging quality, therefore the purpose is summative.
5. Assessment is concerned with process, while evaluation focuses on product.
6. In an assessment, the feedback is based on observation and positive & negative points. In contrast to evaluation, in which the feedback relies on the level of quality as per set standard.
7. In an assessment, the relationship between assessor and assessee is reflective, i.e. the criteria are defined internally. On the contrary, the evaluator and evaluatee share a prescriptive relationship, wherein the standards are imposed externally.
8. The criteria for assessment are set by both the parties jointly. As opposed to evaluation, wherein the criteria are set by the evaluator.

CONCLUSION

An effective, goal-oriented, teaching-learning sequence contains clearly understood objectives, productive classroom activities, and a sufficient amount of feedback to make students aware of the strengths and weaknesses of their performances. Assessment and evaluation are related to both instructional objectives and classroom learning activities and are indispensable elements in the learning process. They are useful for gathering data/information needed into various interests. The data can be used to make decision about the content and methods of instruction, to make decisions about classrooms climate, to help communicate what is important, and to assign grades. Among other techniques to do evaluation and assessment, The teachers can use tests to evaluating and assessing, starting from the small one, incorporating evaluation into the class routine, setting up an easy and efficient record-keeping system, establishing an evaluation plan, and personalizing the evaluation plan.

REFERENCES

1. Bachman, L. F. (1995). Fundamental considerations in language testing. Oxford: Oxford University Press.
2. Brown, D. H. (1990). Language assessment: Principles and classroom practices. London: Longman
3. Darling-Hammond, L. (2006). Assessing teacher education: The usefulness of multiple measures for assessing program outcomes. *Journal of Teacher Education*, *57*(2), 120-138.

4. Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: an advanced resource book.* London ; New York: Routledge.

5. Kellaghan, T., & Stufflebean, D.L. (Eds) (2003). International Handbook of educational evaluation. Dordrecht: Klüver Academic Publisher

6. Kizlik, B. (2010). How to Write an Assessment Based on a Behaviorally Stated Objective. [online Document] Available at http://www.adprima.com/assessment.htm Accessed on September 15, 2017.

7. McAlpine, M. (2002). Principles of Assessment. Glasgow: University of Luton. Available at http://caacentre.lboro.ac.uk/dldocs/Bluepaper1.pdf

8. Merriam-Webster's collegiate dictionary (11th ed.). (2017). New York, NY: Merriam-Webster.

9. Richards, C. and Schmidt 2002. *Longman Dictionary of Language Teaching and Applied Linguistics.* (3rd edition). Harlow, Essex: Pearson Education.

10. Shohamy, E. (1993). The Power of Tests. The Impact of Language Tests on Teaching and Learning. Washington, DC: NFLC Occasional Papers..

11. Shohamy, E. (2001). *The Power of Tests: A Critical Perspective on the Uses of Language Tests.* Harlow: Pearson Education.

12. Weir, J. C. (2005). *Language testing and validation: Evidence-based approach.* New York, NY: Palgrave Macmillan.

13. Weir, J. C., & Roberts, J. (1994). *Evaluation in ELT.* Oxford: Blackwell