UNIT 1. Test and Assessment: The Definitions

*[Banguis]*

A**.** What is a test**?**

1. Something (such as a series of questions or exercises) for measuring the skill, knowledge, intelligence, capacities, or aptitudes of an individual or group.

2. A procedure, reaction or reagent used to identify or characterize a substance or constituent.

Example: written and driving exams given before issuing a driver’s license.

B. Assessment and Teaching (Formal vs. Informal Assessment)

* **Formal Assessment-** have data which support the conclusions made from the test. We usually refer to these types of tests as standardized measures.
* **Informal Assessment-** are not data driven but rather content and performance driven.

For example: running records are informal assessment because they indicate how well a student is reading a specific book.

OVERALL: The assessment used need to match the purpose of assessing. Formal or standardized measures should be used to assess overall achievement, to compare a student’s performance with others at their age or grade or to identify comparable strengths and weaknesses with peers.

**[Now let’s move on to formative vs summative]**

* **Formative –** to monitor student learning and provide ongoing feedback to staff and students. Formative assessment can be tutor led, peer or self-assessment. Formative assessments have low stakes and usually carry no grade, which in some instances may discourage the students from doing the task or fully engaging with it.

**Explanation**

* It is assessment for learning
* It helps student identify their strengths and weaknesses
* Can enable students to improve their self-regulatory skills so that they manage their education in a less haphazard fashion that is commonly found.
* **Summative-** to evaluate student learning at the end of an instructional unit by comparing it against some standard or benchmark.

**Explanation**

* Often have high stakes and are treated by the students as the priority over the formative assessment. However, feedback from summative assessment can be used formatively by both students and faculty to guide their efforts and activities in subsequent courses.

*[Abacahin]*

C. Norm-referenced and Criterion-referenced

**Norm-referenced**

* Norm-referenced assessment is a type of standardized test that compares students’ performances to one another. Norm-referenced assessments compare a student’s performance to the course median.
* Norm-referenced assessment refers to an assessment that ranks students on a “bell curve” to determine the highest and lowest-performing students.
* This method is used to understand how students’ scores compare to a predefined population with similar experience.

**Criterion-referenced**

* Criterion-referenced assessment means that teacher judgements about how a student does in an assessment task are based on standards and criteria that are pre-determined and made available to students at the time the assignment is set.
* **Standards** are a specified and definite level of achievement that may be attained.
* **Criteria** means the characteristics by which the quality of something may be judged.

APPROACHES TO ENGLISH LANGUAGE TESTING: A BRIEF HISTORY

**PRESCIENTIFIC ERA**

In this era, language teachers used some intuitive procedures following the activities of what was later called the “grammar-translation” method of teaching. The tests in this period did not correspond to principles of any theoretical framework because there was simply no theory of teaching or testing at the time. Nor were psychometric principles such as reliability and validity are much in use because these concepts had not yet been implemented in the field. Tests were often made on the basis of teachers’ intuition and judged by them as valid because tests mainly included translations from or into the target language that corresponded to what was practiced in teaching contexts.

**SCIENTIFIC ERA**

With the scientification of language education in the late 1940s and early ’50s, modern language teaching began to enjoy the guidelines offered by scientific disciplines such as linguistics and psychology. Linguistics attempted to describe the nature of language, and psychology began to explain the nature of learning processes. The influence of linguistics and psychology in the field of language education marks the beginning of the scientific era in language testing and teaching. Ever since, most of the developments in language testing have been rooted in the developments in the theories of linguistics, psychology, and other disciplines in applied linguistics such as sociolinguistics, second language acquisition, and others.

**STRUCTURALISM**

In the early 1960s, the dominance of structuralism in linguistics and behaviorism in the psychology of learning led to the emergence of the most popular method of language teaching referred to as the “audiolingual method”. Following the principles of this method, Lado (1961) offered the first linguistically oriented theoretical framework of language ability. According to his model, language consisted of sounds, words, and sentences manifested in the four language skills. Further, language ability was assumed to be the sum of the knowledge of an individual on these components and skills.

Lado’s framework implied that language tests should target measuring these components, assuming that the sum of the performance on these components would be an indication of the learners’ overall language ability. The proponents of this framework began to criticize the tests developed and used in the intuitive era. The major criticism was the lack of objectivity in scoring that implied unreliability of those tests. Another criticism was the lack of clarity of the construct being measured questioning the validity of these tests.

The tests in this era were designed to measure language ability and were supposedly made up of as many elements of language components and skills as possible. Along with modifications in the content of the tests, attempts were also made to establish psychometric characteristics of the tests such as reliability and validity. To establish reliability, one of the main requirements was objectivity in scoring and one way to achieve this goal was to develop multiple-choice items. In addition, to claim validity, the test content had to be representative of the content of the materials to be tested. To achieve representativeness, the tests needed to include as many elements of language as possible.

**Integrative Sociolinguistic Era**

Carroll’s suggestion triggered a new era in language testing and was popularized by Oller in the 1960s and ’70s as integrative, and later as pragmatic, and was referred to as the “integrative sociolinguistic era” by Spolsky (1978). Adhering to Gestalt psychology that the whole is different from the sum of the parts, Oller’s main argument was that the sum of the scores on discrete point items would not add up to a reasonable indication of test takers’ overall language ability. He further believed that the contribution of discrete point items of language to the overall language ability is neither clearly identifiable nor significant. Therefore, he hypothesized that language ability is unitary in nature and all language components and skills are different manifestations of this ability. In contrast to discrete point tests, then, the development and use of integrative tests such as cloze and dictation became popular testing techniques (Oller, 1983).

Oller’s so-called “unitary factor hypothesis” generated extensive research in the field. To support the unitary or indivisible factor hypothesis, Oller and his associates popularized the use of statistical techniques, especially factor analysis, in analyzing test data. The majority of researchers following this hypothesis reported strong empirical support for the unitary language ability hypothesis. They even pushed it further to claim that all mental abilities such as aptitude, attitude, and even intelligence, are manifestations of this unitary language ability. This extreme position that general language ability is an indication of almost all mental abilities was the focus of research for almost a decade.

Oller’s findings along with his strong claims motivated researchers to employ alternative statistical analyses to investigate the validity of his claims. The major objection was the accuracy of analyses and the appropriacy of the interpretations of the statistical findings. For instance, Bachman and Palmer (1982) applied multitraitmultimethod analysis, and Farhady (1983) utilized different versions of factor analysis and both came up with findings different from those reported by Oller and his followers. The outcome of the reanalysis of several sets of data led researchers to believe that the unitary hypothesis was mainly the outcome of the artifact of inaccurate application and misinterpretation of the findings of factor analytic techniques. According to these researchers, language ability consisted of multiple underlying factors though a single ability was found to be common to all mental tests including language tests.

**COMMUNICATIVE ERA**

At the peak of arguments on the divisibility or indivisibility hypothesis regarding the nature of language ability, the field of language testing witnessed another breakthrough with the emergence of the communicative era in the late 1970s. Canale and Swain (1980) initiated this movement by offering a new theoretical framework for language teaching and testing. A significant development in this era was that linguistic competence, which was the focus of defining and measuring language ability, was considered inadequate for defining and insufficient for measuring language ability.

According to this framework, language ability or communicative competence, as they said, consisted of three components including linguistic competence that accounted for what both Lado and Oller considered to be: (a) language ability, (b) social or discourse competence, and (c) strategic competence. In fact, in addition to the linguistically oriented framework of Lado and the psycholinguistic framework of Oller, discourse competence and strategic competence were added to the definition of language ability.

This model was a remarkable advancement in conceptualizing the nature of the language ability construct and has been quite influential in forming later frameworks. Almost all theoretical and operational treatments of language ability that was introduced later had a trace of Canale and Swain’s model. For instance, Farhady (1983) attempted to improve the model by stating that it was linear in structure and additive in nature meaning that communicative competence would be the sum of other competencies. He also argued that communicative competence is so vast in domain and complex in nature that even native speakers may not achieve all its dimensions. He suggested an interactive model of language ability and called it the “model of functional competence” after the functional notional method of teaching.

**Pedagogical Implications**

A major problem with the above-mentioned models, no matter how elaborate they might be, is that they do not translate into the working knowledge of the practitioners. Except for some professional organizations such as ETS, IELTS, Michigan, and recently Pearson, classroom teachers, who are the practicing agents of the theoretical advancements were neither professionally equipped to interpret the complex models of language ability nor could they implement the theories in practice. In addition, a few movements in education and related areas, beginning in the 1990s, led to significant changes in the conceptualization of language testing that put teachers and instruction in the spotlight.

First, the educational reformists in the USA capitalized on the major dissatisfaction with the standard tests that they had little or no relevance to what actually was going on in a real authentic class where real learning happened. Second, in contrast to the priority of reliability in dealing with psychometric properties of a test, Messick’s (1995) reformulation of the progressive pattern of the concept of validity led to the priority of validity over reliability. Finally, his claim of consequential validity brought a whole bunch of variables such as ethics, fairness, authenticity, and use and misuse of test power into the testing picture.

These movements resulted in a drastic change in approaches to measuring and evaluating learners. They forced the field of language assessment to shift attention from modeling language ability to finding alternatives to assessing language ability in the context of learning. Therefore, a silent division was made between standard proficiency tests measuring overall language ability, developed and administered by professional organizations on the one hand, and those of curriculum-based and teacher-made tests on the other. Scholars attempted alternative assessments that would be in line with and integrated into the process of learn ing. The concepts of assessment for learning and during learning became popular to replace or at least overshadow assessment of learning that was the main purpose of standard general proficiency tests.

*[Ayop]*

D. Current Issues in Classroom Testing

Student assessment is a critical aspect of the teaching and learning process. It enables teachers to measure the effectiveness of teaching with student performance to specific learning objectives. According to Ebel (1972), the increase in the use of educational tests has been accompanied by an increase in criticism of the practice. Tests vary in quality, with some being particularly poor. He argued that educational testing may be socially detrimental for a number of reasons. In the context of educational practice there are some crucial issues with respect to assessment, which demand the attention of all those who are concerned with the quality of teaching and learning process.

Here are some:

* **Poor Test Quality**

Tests may not show sufficient evidence of validity and reliability. Many tests used for educational assessments are not standardized or prepared not by undergoing the systematic test development procedure and applying psychometric principles. Such test does not possess the good qualities and fails to perform their functions and purpose. For example, the question papers used by many universities are criticized on the ground that they are inferior in quality and they fail to perform their educational functions of assessing and evaluating (Areekkuzhiyil, 2019). One of the reason for the poor quality of test are that the construction of good quality test require expertise, it is time consuming process and has to undergo a series of sequential procedures.

* **Domain Dependency Issue**

Cognitive scientific research reveals that general and specialised knowledge function in close partnership. To be maximally effective, assessment requires the interaction of general principles, strategies, and techniques with reasonably deep cognitive domain understanding. That deep cognitive domain understanding includes the processes, strategies and knowledge important for proficiency in a domain, the habits of mind that characterize the community of practice in that domain, and the features of tasks that engage those elements. It also includes those specialized aspects of domain knowledge central to helping students learn. A teacher who has weak cognitive domain understanding is less likely to know what questions to ask of students, what to look for in their performance, what inferences to make from that performance about student knowledge, and what actions to take to adjust instruction. The intellectual tools and instrumentation given to teachers may differ significantly from one domain to the next because they ought to be specifically tuned for the domain in question.

A possible approach to dealing with the domain dependency issue is to conceptualise and instantiate formative assessment within the context of specific domains.

* **Measurement Issues**

Educational measurement involves four activities: (i) designing opportunities to gather evidence, (ii) collecting evidence, (iii) interpreting it, and (iv) acting on interpretations. Assessment is not simply the elicitation of evidence but also includes making inferences from that evidence. Assessment is an *inferential* process because others cannot know with certainty what understanding exists inside a student’s head. They can only make conjectures or hypotheses based on what we observe from such things as class participation, class work, homework, and test performance. The measurement issue lies in the interpretation of evidences for learner performance and achievement. For example a weak performance in mathematics may be due to linguistic deficiency, but the same would wrongly interpret as underachievement in mathematics. This misinterpretation would lead to unnecessary course of action.

* **System Issues**

It refers to the fact that assessment exists within a larger educational context. If that context is to function effectively in educating students, its components must be coherent. Gitomer and Duschl (2007) describe two types of coherence, internal and external. Assessment components can be considered internally coherent when they are mutually supportive; in other words, formative and summative assessments need to be aligned with one another. Those components must also be externally coherent in the sense that formative and summative assessments are consistent with accepted theories of learning, as well as with socially valued learning outcomes. External coherence, of course, also applies to other system components like the educational ideology, policies and programmes. In any event, if these two types of coherence are not present, components of the system will either work against one another or work against larger societal goals. Thus, the effectiveness of assessment will be limited by the nature of the larger system in which it is embedded and, particularly, by the content, format, and design. Ultimately, we have to change the system, not just the approach to assessment, if we want to have maximum impact on learning and instruction. Changing the system is a very big challenge indeed.

* **Bias**

Blatchford and Cline (1992) suggest that the assessment process should operate without bias with respect to gender, social class, ethnicity, language use and religion. However, many researchers have demonstrated the existence of bias in educational assessment. Many researchers have argued that most assessments are culture biased and discriminates against certain ethnic groups. Anastasi (1972) argues that it is not productive to attempt to develop tests that are ‘culture-free’ (free from cultural influences) and, instead, there should be efforts to develop tests that are ‘culture-fair’ (common to different cultures). An example of bias would be a reading comprehension test that uses a reading passage that reflects a middle class, white family. For many people, such an experience is unknown for them. When they try to answer the questions they lack the contextual knowledge of someone who is familiar with this kind of situation and this puts outsiders at a disadvantage.

E. Critical Language Pedagogy

Critical language testing (CLT) is a philosophical approach that states that there is widespread bias in language testing. This view is derived from critical pedagogy, which views education as a process manipulated by those in power.

There are many criticisms that CLT has of language testing such as the following.

* Test are deeply influenced by the culture of the test makers
* There is  a political dimension to tests
* Tests should provide various modes of performance because of the diversity in how students learn.

**Testing and Culture**

CLT claim that tests are influenced by the culture of the test-makers. This puts people from other cultures at a disadvantage when taking the test.

**Politics and Testing**

Politics and testing is closely related to the prior topic of culture. CLT claims that testing can be used to support the agenda of those who made the test. For example, those in power can make a test that those who are not in power cannot pass. This allows those in power to maintain their hegemony. An example of this would be the literacy test that African Americans were required to pass in order to vote. Since most African Americans could not read they were legally denied the right to vote. This is language testing being used to suppress a minority group.

**Various Modes of Assessment**

CLT also claims that there should be various modes of assessing. This critique comes from the known fact that not all students do well in traditional testing modes. Furthermore, it is also well-documented that students have multiple intelligences.

It is hard to refute the claim for diverse testing methods. The primary problem is the practicality of such a request. Various assessment methods are normally impractical but they also affect the validity of the assessment. Again, most of the time testing works and it hard to make exceptions.

**Conclusion**

CLT provides an important perspective on the use of assessment in language teaching. These concerns should be in the minds of test makers as they try to continue to improve how they develop assessments. This holds true even if the concerns of CLT cannot be addressed.

Prepared by:

Banguis, Lesly

Abacahin, Sol Krishna

Ayop, Jane Irish E.

*References:*

Areekkuzhiyil, Santhosh. (2021). Issues and Concerns in Classroom Assessment Practices. *Edutracks*, 20(8), pp 20-23.

Ebel, Robert L. (1972). *Essentials of Educational Measurement*. N.J.: Prentice-Hall, Inc.

Educational Research Technique. Critical Language Testing. <https://educationalresearchtechniques.com/2017/06/26/critical-language-testing/#:~:text=Critical%20language%20testing%20(CLT)%20is,manipulated%20by%20those%20in%20power>.

Formal and Informal Assessment. <https://www.spriglearning.com/formal-formative-assessment-or-informal-formative-assessment-whats-right-for-early-learning/>

Fulcher, G., Davidson, F. (2007). Language Testing and Assessment An Advanced Resource Book. Routledge.

<https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118784235.eelt0343#:~:text=Abstract,learner%20understood%20what%20was%20taught.&text=Since%20then%20different%20conceptualizations%20of,learned%20language%20has%20been%20measured>

<https://tophat.com/glossary/n/norm-referenced-assessment/>

<https://uoa.custhelp.com/app/answers/detail/a_id/15450/~/criterion-referenced-assessment>

Test and Assessment. <https://www.definitions.net/definition/test>