



Need for Understanding Theoretical Underpinnings of Testing Language Skills: Basic Concepts for Reliable Practice

Dr. Abdul Latheef Vennakkadan
Head, Department of English
Government College, Kondotty
Kerala, India

Abstract

Testing occupies an unassailable logic in the teaching and learning paradigm as it continues to be the most reliable source of data to give strategic insights to the cognitive and performative competencies of learners, functional efficacy of instructional practices and in making evaluative decisions on the quality of an educational system. Tests, specifically in the context of assessing language skills, have to be valid, authentic and must be exploratory enough to indicate differences in the abilities of learners. The two major uses of language tests are as sources of information for making decisions within the context of educational programs and as indicators of abilities of attributes that are of interest in research on language, language acquisition and language teaching. A language test designed without a reasonably profound know-how of the theoretical underpinnings and the must-have attributes of an ideal test, instead of providing authentic data, can lead into contagious issues of far-reaching social and educational consequences. So the paper attempts a brief survey of unavoidable concepts and attributes that any language test designer should be cognizant with for a close alignment of testing and instruction.

Key Words

Assessment, Evaluation, Diagnostic/Proficiency tests, Test Attributes, and Validity

Introducing Concepts in Testing

The history of language testing and assessment can be traced back to the origin of teaching and learning. The role of tests or instruments of assessment in language education is to capture the kind of behaviour or performance of learners which can act as reliable evidence for their skills in the Target Language (TL) in real life contexts. The tools used for evaluation have to be reliable, accurate and must be sensitive enough to indicate differences in the abilities of



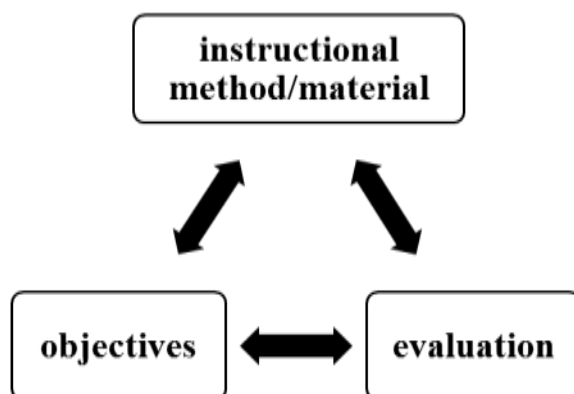
learners. Testing, Measurement, Evaluation and Assessment are oft-encountered terms in the literatures on language testing. Though mutually inclusive and interrelated, a small distinction can be observed between/among these terms when a close and subtle analysis is attempted.

On the one hand, an examination is quite formal and has an element of finality in it. On the other hand, a test is informal and is an on-going activity along with classroom instruction, and is of lesser duration. Tests are tools used to measure abilities or performance of learners. Carroll (1968) describes a test as “a psychological or educational procedure designed to elicit certain behaviour from which one can make inferences about certain characteristics of an individual” (p. 46).

Evaluation, as the term indicates, is broader in scope and encompasses the entire process of gathering information and making conclusive inferences after a judicious analysis of the obtained details. Measurement is said to be the activity of assigning numerical values to the obtained data in a meaningful, consistent and scientific manner. The process of estimating abilities of learners in the language using the marks, scores or grades secured in different tools of testing is termed as assessment. For the sake of consistency and uniformity, the term testing has been used in the study as an umbrella term to mean all the related terms of language assessment.

The interrelationship between testing, teaching and learning, methods and materials is inseparably interwoven. The objectives of teaching determine the type of material and methods to be adopted and all these factors in conjunction decide the assessment techniques to be employed. The diagram below throws clearer light on this interrelationship. The arrows point in both directions to underscore the fact that their interlink is not one-sided but a two way relationship.

Figure. 1. Interlink between evaluation, instruction and objectives





The two major uses of language tests are as sources of information for making decisions within the context of educational programs and as indicators of abilities of attributes that are of interest in research on language, language acquisition and language teaching (Heaton, 1975, p.36). The researches on impact studies (Bachman and Palmer, 1996) underpin the fact that both testing and teaching are so closely interrelated that it is virtually impossible to work in either field without being constantly concerned with the other. Tests serve different purposes like reinforcing learning, motivating students, a means of assessing the students' performance in the language. A good classroom test helps locate the precise areas of difficulty encountered by the class or by the individual student and their areas of strength as well. Thus, test is used as a valuable teaching as well as assessment device.

When speaking about the relationship between testing and teaching, Hubbard et al. (1983, p. 255) mentioned ten aims that testing fulfils.

1. Identification of problem areas for remedial attention.
2. Giving each student a course grade.
3. Assessment of your own effectiveness as a teacher.
4. Checking on general progress and obtaining feedback.
5. Course or syllabus evaluation.
6. Preparation for public examinations.
7. Institutional requirement for student promotion.
8. Measuring what a student knows.
9. Identification of levels for later group-work.
10. Reinforcement of learning and student motivation.

Types of Assessment

Based on the focus and purpose of the assessment, it is broadly divided into two, namely formative and summative. Though they are dissimilar in intension and function, at times, the same type can meet purposes of both kinds. To ensure comprehensive and reliable evaluation of the abilities of students, it is better to use both types in assessment practices.

Formative Assessment

Formative assessment primarily aims at monitoring students' learning and providing on-going feedback for teachers to review or revise their teaching and for students to improve their learning. It is conducted during a course or programme and is otherwise known as internal





tests, seasonal tests or progress tests. These types of tests are often teacher made, teacher evaluated and are effective means to understand whether the set objectives are being realized or the mode of instruction and teachings materials need a review.

More specifically, formative assessments help students identify their strengths and weaknesses and target areas that need improve further. It helps the teachers recognize where students are struggling and address problems immediately .Formative assessments are generally low stakes and help in giving the teacher a record of learner’s performance and progress in each stage of learning or skill acquisition. Examples of formative assessments include asking students to draw a concept map in class to represent their understanding of a topic, asking students to answer a few questions of objective type at the end of a class or instructing students to write a summary of the topics they learned in the previous class.

Summative Assessment

This type of assessment is held at the end of a course of study or a programme to evaluate student learning by comparing it against some standards or benchmarks. This kind of test is high stakes because they have high value and is used for selection or certification. They are more formal than formative tests, generally based on a syllabus and are conducted by an external authority or a Board of Examination. The main purposes of summative assessment are to provide feedback about the learner’s performance, level of achievement at a particular point of time or completing some specific units in a course of study. This being a sort of achievement test, the learners’ scores are taken as indicators of their mastery of skills in language or attainment of the targeted learning outcome of the course. Examples of summative assessments include midterm exams, and term-end exams. Information from summative assessments can be used formatively when students or faculty use it to guide their efforts and activities in subsequent courses.

Types of Tests

Tests are means of obtaining systematic evidence on which instructional decisions can be based. Educators see tests as motivators that stimulate individuals to do their best. If they are well designed and properly used, tests can effectively enhance the educational process (Richards, 1990). Educational testing is in fact a world endeavor. In everyday life too, the act of testing is going on, knowingly or unknowingly to select an object, to assess a person so on. Therefore, tests are inevitable tools in educational practices to provide information about the achievement of the testees, without which inferences can’t be drawn nor decision can be



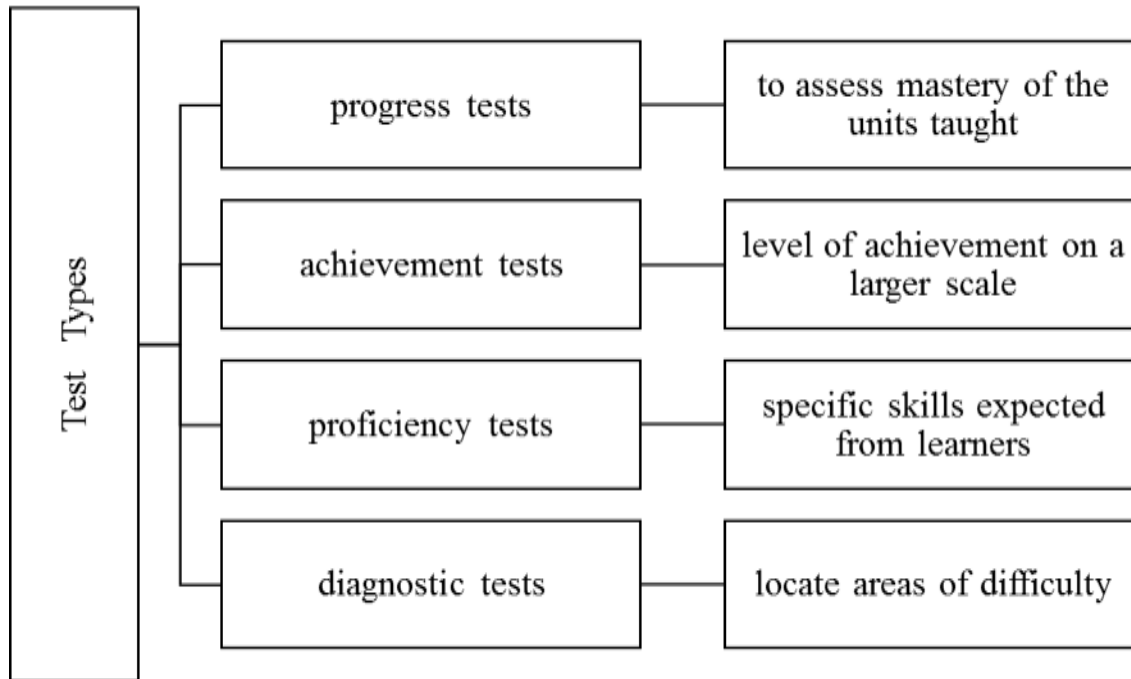


taken. Depending upon the purpose for which a test is designed, they can be compartmentalized into different kinds.

Progress Tests are those which are often constructed by the teachers and are designed to measure the extent to which students have mastered the material taught in classrooms. They are conducted, after teaching each prescribed modules, to assess teachers' work and students' learning. It is a teaching device to reflect feedbacks on the efficacy of instructional mode, materials and student achievement. An achievement test, as the name reveals, is meant to measure how far the learners have achieved what is taught to them. It is similar to PT in purpose but is far more formal and is intended to measure achievements on a larger scale. ATs are often written based on what a student is presumed to have learnt, not necessarily on what they have actually learnt or been taught. Its value and importance is much more than PT as scores in AT are significant for certification or promotion. Proficiency Tests, in contrast, look forward to what the students will be required to perform at a particular level of education. They are no way related to any syllabus or teaching program because they look into level of skills expected at certain stages. It is concerned with measuring students' control of the language in the light of what they will be expected to do with it in their future performance of a particular task. It is meant to measure not general attainment but specific skills that the students may require in their future career/study. Diagnostic tests are always administered to determine/locate the areas of difficulty of learners and to devise appropriate remedial measures. It is held during the course of study so that both teachers and learners get a good picture of those areas where they need to focus and improve further. They are formative in nature and are low stake tests. Diagnostic Tests are primarily designed to assess students' knowledge and skills in particular areas before a course of study is begun. The major test types and their purposes can graphically be illustrated using the diagram below:

Figure .2: Test Types and Purposes





Attributes of an Ideal Test

The credibility and effectiveness of assessment depend on the validity of the instruments or tools used to measure what is to be measured. Bachman and Palmer (1999), while commenting on designing and developing useful language tests, observe that tests differ from other components of instructional programme like teaching material, learning activities for their primary purpose of measurement. Fulcher and Davidson (2007) observe “there is something fundamentally different about language testing. Language testing is all about building better tests, researching how to build better tests and, in so doing, understanding better the things that we test. Sociolinguists do not create ‘sociolinguistic things’. Discourse analysts do not create discourses. Phonologists do not create spoken utterances. Language testing, in contrast, is about doing. It is about creating tests”.

Of all the qualities of an ideal language test - reliability, validity, authenticity, discrimination, positive washback and practicality- the first two are critical for tests because they are the qualities that provide the major justification for using test scores –numbers as a basis for making inferences or decisions. These essential requisites can be elaborated as follows.

Reliability





In simple terms, it is defined as consistency of measurement which helps yield consistent test scores across different characteristics of testing situations. The notion of reliability implies that if the same tests were to be administered to the same test takers on two different occasions, in two different settings, the test scores in both the occasions have to be justifiably similar. In other words, students who obtain high scores on one set of items also obtain high scores on other sets of equivalent items, and those who have a low score on one set of items also have a low score on other sets of items (Scannel and Tracy, 1975). Thus, reliability is the degree of stability and trustworthiness of the instrument which has been used to gather data. Otherwise reliability refers to a particular kind of consistency of results obtained over a period of time, over different groups but using similar tests (equivalence) and over different parts of a test.

Validity

This refers to how far a test includes indicators of constructs (abilities) of the test takers' language competence that a test is expected to measure. The term, thus, is used to mean the extent to which scores on a given test can be interpreted as evidences of the abilities we intend to measure. Validity is established not by declaration but by evidence. To phrase it another way, validity in testing refers to whether the test measures what it claims to measure, and whether it measures what was taught. For example, a test which is designed to determine the extent to which a particular group of students have mastered specific algebraic concepts will not be valid when administered to a different group of students with the intent to determine their performance in Elizabethan literature. Similarly, a test of English as a Second Language (ESL) is not valid for students learning translation theory (Heaton, 1995). Messick's (1995) unified view of validity predicated that validity is a multifaceted concept, which can only be established by integrating considerations of content, criteria, and consequences into a comprehensive framework for empirically testing rational hypotheses about score meaning and utility.

Authenticity

It is another equally important feature of a good test. In communicative testing, authenticity is a key element in the designing of materials and test items. It means assessing language behaviour by observing it in real, or at least realistic, language-use situations which should be as authentic as possible (Gronlund, 1985).

Discrimination

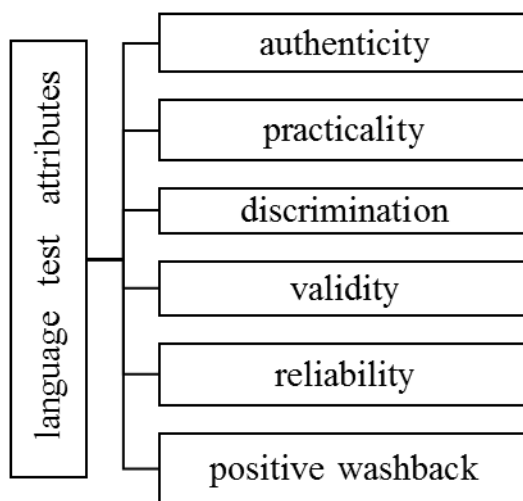


This feature an ideal test implies the capacity of a test to differentiate among the different candidates and to reflect the differences in performances of the individuals in the group. For example, a test in which all the testees score more than 70% clearly fail to discriminate the more talented and the less talented among the candidates. It is dependent on the degree of difficulty of the test items.

Positive Washback

Washback is defined as the influence of testing on teaching and learning process. If a test is said to be harmfully influencing the scope and objectives of a course, limiting teaching to training for tests, it is said to be negative washback. Positive washback is said to exist, if the testing strategies promote and enhance the course objectives in a beneficial manner. An ideal test should have a positive washback on teaching, learning and language acquisition. Graphically, these attributes can be better illustrated as follows:

Figure-3. Attributes of in Ideal Test



Conclusion

In sum, testing has been an indispensable corollary to any kind of formal educational practices which are often critiqued, reviewed, reinforced and reformed based on the data obtained from testing. In other words, both teaching and testing are interrelated in such a way that the limitations /strengths of one can be overtly reflective on the other. An adequately erudite knowledge of the theoretical constructs of writing an ideal test to assess language skills of





learners and how to incorporate them in the form of questions play a pivotal role in deciding success of any language program offered. So, a sound idea of testing and test writing should be part of the core components of programmes offered at teacher training level.

References

Bachman, L.F., Palmer, A.S.,(1996). Language Testing in Practice. Oxford University Cambridge: CUP

Carroll, BJ.(1968) Specifications for an English Language Testing Service. J Charles Alderson and Arthur Hughes (Eds) .Issues in Language Testing. BBC. London

Fulcher, G and Davidson, F (2007) Language Testing and Assessment. Routledge, London.

Gronlund, N. E., (1985). Measurement and Evaluation in Teaching. NY: Macmillan.

Heaton, J. B., (1995). Writing English Language Tests. London: Longman Group.

Hubbard, Peter & et al (1983). A Training course for TEFL .OUP. Oxford.

Richards J C & T S Rodgers (1986) .Approaches and Methods in Language Teaching.

