

# LANGUAGE TESTING : ITS PLACE IN THE TEACHING PROCESS



\* Asst.Prof. Dr. Vinit Phinit-Akson

Prior to tackling the task of testing, it is appropriate that one should understand the role of testing--its place in the teaching process. Educational psychologists use a special model for describing the inter-relationship between testing and the other related components of the teaching process. More specifically, educational psychologists speak of

*a basic teaching model.*

## BASIC TEACHING MODEL

The basic teaching model (BTM) consists of four main components: instructional objectives; entering behavior; instructional procedures; and performance assessment. This is illustrated in the following diagram.

\* Dean, Faculty of Humanities, Dhurakijpundit University : Ph.D. (Linguistics) University of Pittsburgh, U.S.A.

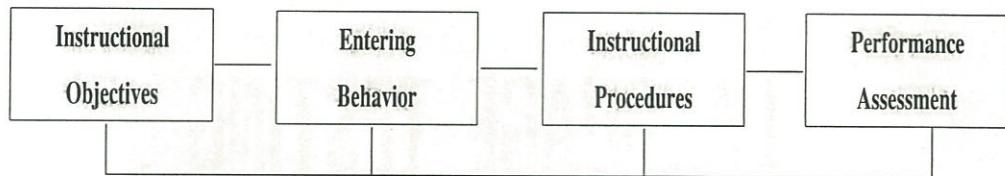


Figure 1.1 BTM with Feedback Loops.

Feedback loops are those lines which connect components later in the sequence with earlier ones. In the above figure, the three feedback loops connect performance assessment (testing) with the preceding components of the basic teaching model. In order to understand the model better, we shall look more closely at each component.

Instructional objectives are those objectives the student should attain after completion of a section of a course of instruction. These objectives can vary in scope from the ability to reproduce a few simple dialogues of a conversation to that of the ability to speak and comprehend a target language (English as a second language--for example).

Entering behavior is a special terminology used to refer to the student's level of knowledge (in a particular field) before instruction has begun. It refers to what the student has learned previously, his intellectual ability and development, his state of motivation, and the social and cultural determinants of his learning ability. This special

terminology (entering behavior) is a more accurate name than its usual alternatives (readiness, human ability, and individual differences). It should also be noted that although the BTM places instructional objectives before the student's entering behavior, these two components interact in practice.

Instructional procedures describe the actual teaching process. This consists of the tests and observations used in determining how well the student has achieved the instructional objectives set for that particular course of instruction. Should such performance assessment show that the student has fallen short of mastery or some lesser standard of achievement, then one or all of the preceding components of the BTM may require adjustment.

#### Types of Language Tests and Their Usages

In language testing, there is often confusion as to the terminology employed in relation to the various types of language tests. However, one can classify such tests according to their purpose. Thus, the terms *aptitude test* and *proficiency test* refer to tests which

measure the students' present ability for future learning. Specifically, the aptitude test measures the examinee's suitability for a particular field (for example, language learning).

The *proficiency* test is used to find out whether the examinee has mastered specifics skills and content considered necessary for a particular course of study. Sometimes, such a test is called a *placement test*. This is a test to assign groups of incoming students with different types of preparation to different levels of classes.

*Achievement or attainment tests* have a different function from proficiency tests. Achievement tests are used to measure what has been learned of a certain syllabus or course. An example of an achievement test is the teacher made test given at the end of a course (or at the middle such as a mid-term test) of instruction to measure the terminal behavior (knowledge gained in the course over and above the level of knowledge at the start of the course eg. entering behavior). It is from such measurements of terminal behavior that students are given grades.

The next type of language test that the teacher should know about is the *diagnostic test*. Perhaps we can understand such a test better if we compare it with achievement and proficiency tests. According to E. Ingram (1968, p. 73), achievement and

proficiency tests answer the question "How much does the learner know?" Diagnostic tests answer the question "What does the learner know?" Moreover, apart from answering the "what question", a diagnostic test should also be able to answer the "why question". Thus, diagnostic tests give quantitative assessments and qualitative assessments. They give a detailed picture of what the learner is good or bad at. One example of a diagnostic test is the diagnostic test for reading. In diagnostic testing, the concept of relevant behavior must be analyzed in greater detail than in achievement and proficiency testing since the purpose in diagnostic testing is normally to get information for remedial teaching.

#### Exam Techniques (Types of Tests Used)

**Their Strengths and Weaknesses.**  
There are several ways to test language behavior. Some techniques are very well known as they have been used for a very long time. Others are of a more recent origin. The following, lists some exam techniques used. Each technique has its own strength and weakness.

**1. Translation** as a testing device, has been used very often both in the past and at present. It has also been used as a teaching device by the so-called traditional school of language teaching. However, modern linguists (especially those trained in structural

linguistics) are not in favor of translation- neither as a testing device nor as a teaching technique. Some linguists even go so far as to say that translation can impede the process of language learning. Moreover, since various translation are possible, who is to judge which is the best one? Thus, evaluation of any translation is rather difficult and is bound to involve some kind of value judgement or other.

2. **Dictation** has been used as a testing device to assess student performance in the phonological, grammatical, and lexical areas of his or her language behavior. Dictation has some relevance for beginning and intermediate levels of English. However, it is time consuming to conduct and is tedious to correct. Thus, other types of tests are more efficient.

3. **Essay tests** or composition tests allow the examinee to use his originality and compose his own written response to test problems set up by the teachers. Such essay tests range from short paragraphs to long essays. The examinee may be expected to demonstrate not only his use of grammar and vocabulary but also his ability to organize ideas. When one desires to give scope to the examinee for demonstrating such abilities, essay tests are more suitable than objective types of tests. However, essay tests take time to correct and can be rather subjective.

4. **Oral interviews** are similar to the essay test in many ways but measures oral competency rather than competency in the written language. However, different techniques can be used to measure the various levels of oral proficiency.

5. **Objective tests** or multiple-choice items were developed to overcome certain weaknesses of the essay tests. Objective tests are of a more recent origin. There are a great number of items. Thus, it is not possible for examiners to evade difficult questions or issues. Although they take a longer time to compose, objective tests are quicker to score. Moreover, objective tests have a greater reliability than essay tests. Note, however, that the best approach in testing is to combine both objective and essay types of items in the one test.

**Short-answer items** are those items which are similar to both objective and essay types of items. The answers are, as the name states, brief. Thus, the examinee might be required to either complete a sentence or compose one of his own. Examples are those items which require the examinee to rewrite declarative sentences as negative ones or statements into questions.

#### **The Components of Language and Language Skills**

Human language can be classified into two forms: a spoken form and a written

one. Testing reflects the approach used in teaching at that particular time. Because the old approach to language teaching (traditional approach) emphasized the written form of language, the testing also concentrated on the written aspects of language. Similarly, since linguists and language teachers today concentrate on the spoken form of the language, modern language testing reflects this approach.

In any case, we are dealing with two linguistic activities in both speech and writing. Such activities are referred to as an encoding process (speaking or writing) and decoding process (listening or reading). The reader is referred to the section below on the testing of the listening skill for a more detailed discussion of the encoding and decoding processes.

Linguists interpret language learning

as involving four skills: listening speaking, reading and writing. Language itself is interpreted as having the following components: phonological component, syntactic component and a semantic component. In tagmemic theory, each component is treated as a system. Syntax deals with the grammatical system. The semantic component is taken care of by the lexical component (the vocabulary system). A tagmeme simultaneously has both form and meaning. In language testing then, it is possible (and efficient) to construct tests for testing the various components, namely, phonology, syntax (grammar or structure), and the lexicon (vocabulary). It is possible to diagram the relationship between the language skills and each of the language components as follows:

Components	The Four Language Skills			
	Listening	Speaking	Reading	Writing
Phonology/orthography	phonology	phonology	orthog.	orthog.
Structure	✓	✓	✓	✓
Lexicon	✓	✓	✓	✓

## Test Validity Reliability, Objectivity and Efficiency

A good language test, like other kinds of tests (history test, literature test) is based on four important concepts: validity, reliability, objectivity and efficiency. In this section, we shall be looking at these four important concepts and the relationships between the concepts themselves.

### Validity

A Test is valid (has validity) when it actually measures what it purports to measure. In our basic teaching model (BTM), a test has validity when it actually measures the terminal performance described in the instructional objectives.

The crucial task in constructing a valid test is to get the student to perform (on the test) those things which are required by the instructional objectives.

Since validity has been defined above in terms of both terminal performance and instructional objectives, we are dealing with direct validation. This requires that the instructional objectives are in the form of explicit statements. The teacher can then validate his tests directly by comparing the test items with the instructional objectives.

By defining test validity in terms of instructional objectives, we are indicating that it is more related to absolute rather than to relative standards of performance. This is so

since an explicit statement of instructional objective requires a set minimum standard of acceptable performance. The student who wishes to pass must meet this standard, if not surpass it. If the student falls below this set standard, the exam result is interpreted as a failure. When one is dealing with absolute standards of performance, it is more important for the students to acquire perfect scores (the highest scores possible) than to obtain a distribution of scores. When some students fall below this set standard and fail the course, obviously there is a need for further instruction.

### Reliability

Consistency is the key word here : a reliable test is one which measures terminal performance consistently. Thus, we expect a student to achieve about the same scores on the second administration of a test as he did on the first. A reliable test is one which can be used again and again with consistent results. When a test is unreliable, one can expect erratic or unpredictable results.

According to Ebel (1965, pp. 310-11), three factors account for the degree of test reliability. First, if the test items are too difficult, too easy, or ambiguous, they will produce very unreliable scores. Second, the students's personal behavior (state of alertness, energy, emotional state and so on) must be at about the same level. If such behavior is not

the same at each sitting of the exam, then test reliability will be reduced. Third, the person who scores the test can reduce the reliability factor if he scores the test haphazardly rather than on some standards uniformly applied to all the test papers.

### The Relationship Between Reliability and Validity

Reliability and validity are closely related. First for a test to have validity, it must also have reliability. Therefore, a test which does not measure terminal performance consistently cannot claim to measure anything at all. Second, the opposite of this relationship between reliability and validity is not true. That is, a test can be reliable without being valid. Third, it is possible that by increasing the reliability of the test, the validity is unwittingly reduced. This is due to the fact that some terminal behaviors are more easily measured than others. Good performance assessment, however, demands that the test in fact tests all the learning which is covered by the instructional objectives and not only those types which are more easy to test.

### Objectivity

In an objective test the test scorer is free of subjective judgements. In fact, objectivity is fairly difficult to achieve whether the test questions are written in objective or

essay form. Even with multiple-choice objective types of tests, there is often debate on which response is the best one. With essay types of subjective tests, the responses are even more subjectively scored. The very same scorer may very well award different points for similar answers (to the same question) depending on his or her particular mood at that very moment.

Ebel (1965, p. 296) is of the opinion that when there is general disagreement on the correct answers to objective questions, the trouble can be traced to the exam question itself. Nevertheless, three factors may cause a reduction in the objectivity of various objective items. First, the question itself may have more than one correct answer. Second, the test question may concern opinions and theories about which there is disagreement. Third, the question may be satisfactory when originally conceived but may be poorly worded.

### The Relation Between Objectivity and Reliability

Objectivity and reliability are closely related. If two or more tests are not scored consistently, this can lead to low reliability. Such inconsistency in scoring often occurs in the grading of essay types of questions. Inconsistency can also occur when there is some error in the actual scoring itself.

### Efficiency

"An efficient test is one that gives a great number of independent, scorable responses per unit of time." (Ebel, 1965, 295). Therefore, an efficient test is one that makes the best use of the instructor's and students' time. Whether a particular type of examination is efficient or not depends on the particular circumstance and usage. For instance, an objective type of exam using multiple-choice items is time consuming to prepare. However, it is very efficient to use since the exam can be used over and over again for successive groups. Moreover, such a

test can be quickly scored (either manually or by machine). Nevertheless, an essay type of exam is more efficient when a new exam must be given to many classes and when there is little chance of using such a test over and over again. From the student's point of view, the essay test is not as efficient as the objective test (multiple-choice) since the student must spend most of his time writing rather than reading (students can read ten times as fast as they can write). Here again, there should be some balance between validity and efficiency as well as between validity and reliability. □ □

### REFERENCES

Ebel, Robert L. *Measuring Educational Achievement*. Engelwood Cliffs. N.J. Prentice-Hall, Inc., 1965

Ingram, Elizabeth. *"Attainment and Diagnostic Testing"* in **LANGUAGE TESTING SYMPOSIUM**. London: Oxford University Press, 1968

Vinit Phinit-Akson, *Modern Language Testing (Testing English as a Second Language.)* Bangkok: Thammasat University Press, 1976

Weir, Cyril J. *Communicative Language Testing* New York: Prentice Hall, 1990