

Language Testing & Evaluation

Tazanfal Tehseem

Multan

Resource Person:

1

LANGUAGE IS ...

- The most complex of human behaviors. Like the air we breath – essential, but unnoticed unless it is flawed
- Despite its complexity, inadequate for context-free communication

WHY DO WE *TEACH*?

The *primary* purposes of language *teaching* are to (e.g.):

- promote or facilitate learning;
- enhance learners' linguistic, cognitive, emotional, and social development

WHAT DO WE ASSESS?

- **Learning objectives**
- **“Content” of the syllabus or curriculum**
- **“Content” of lesson plans**
- **“Content” of teaching and learning materials**
- **“Content” of teaching & learning activities**
- **Language ability, proficiency**

WHAT IS A TEST?

- ❖ **A measuring device which we use when we want to compare an individual with other individuals who belong to the same group**
- ❖ **Tests invite candidates to display their knowledge or skills in a concentrated fashion**
- ❖ **A test is a method of measuring a person's ability or knowledge in a given area**
- ❖ **A method that generally requires some performance on the part of either the testee or the tester, or both**

BASIC PRINCIPLES

- **Test what you teach and teach what you test and test on a regular basis**
- **Give feedback quickly to students**
- **Make testing part of the learning process**
- **Make tests reflect the student level of understanding**
- **Never use testing instead of teaching**
- **It is not the only measure for grading**
- **A test should be a sample of what you teach and what learners are trying to learn**

WHY DO WE ASSESS?

Formative decisions:

➤ **Teachers make decisions about:**

- **changing their teaching (materials, activities).**
- **presenting, revising, contextualizing, and scaffolding new material;**
- **placing learners into appropriate groups or levels;**
- **guiding their students' learning;**
- **challenging and motivating their students to learn.**

WHY DO WE ASSESS?

Summative decisions:

➤ **Teachers and administrators make decisions about:**

- which students pass and fail a course.
- which students are certified at a particular level of ability

TEACHERS' QUESTIONS ABOUT ASSESSMENT

Teachers almost always ask:

- *When should I assess?*
- *How often should I assess?*
- *How should I assess?*

Teachers seldom ask:

- *What should I assess?*

Teachers almost never ask:

- *Why should I assess?*

AIMS OF TESTING

Davies: 4 aims for testing

- Research
- Progress
- Guide to teaching and the curriculum
- Representing terminal behavior

Testing is more a means rather than an end.

TESTING IN EFL/ESL

- Test – testing – evaluation – assessment – feedback
a measurement instrument designed to elicit a specific sample of an individual behaviour (gives evidence of the abilities which are of interest)

Language teaching and testing

- the need for a change: in the past they were separated
- a test should reinforce learning / motivate students to learn
- relationship between teaching and testing – *WASHBACK*
- research should support teaching and testing
- influence of standardised tests and public examinations

TESTING IN EFL/ESL

- skill-based teaching - testing
- The four skills:

	PRODUCTIVE	RECEPTIVE
ORAL	SPEAKING	LISTENING
WRITTEN	WRITING	READING

+ VOCABULARY + GRAMMAR – USE OF ENGLISH

- discord between test and understanding of language and communication
- passive recognition of language
- cutoff scores are very problematic
- general proficiency ≠ academic proficiency

TESTING VS. EVALUATION

Tests are conducted in order to *gather information and that* is needed to make decisions. The decision-making process is called **evaluation**.

Evaluation is gathering and judging students' **PERFORMANCES**

In a simple diagram, the relationship between testing and evaluation is:

TESTING → INFORMATION → EVALUATION

- The type of evaluation we have to make determines the kind of information we need.
- In turn, the kind of information we are looking for determines the kind of test we need to conduct:

EVALUATION (determines) INFORMATION (determines) TESTING

So, ultimately, the type of test that we conduct depends on the type of evaluation we need.

THINKING SKILLS IN DESIGNING A TEST

- **What level of learning corresponds to the course content**
- **Bloom's Taxonomy of Educational Objectives**
 - **Knowledge, Comprehension, Application**
 - **Analysis, Synthesis and Evaluation**

TESTS AS DIAGNOSTIC TOOLS

- Students demonstrate learning
- Instructor effectiveness – modify teaching strategies or activities
- Assignment of letter grades
- Measure specific areas of weakness
- Effect the teaching and learning process
- Identify strength and tell what kind of teaching should take place

TYPES OF LANGUAGE TEST (I)

Tests distinguished by use

- Achievement/attainment tests
- Proficiency tests
- Aptitude tests
- Diagnostic tests
- Placement tests

Tests distinguished by the standard for measuring

- Criterion-referenced tests (e.g. achievement tests)
- Norm-referenced tests (e.g. proficiency tests)

**TAKE A
BREAK!!!**

TYPES OF LANGUAGE TESTS

o Achievement tests

- associated with process of instruction
- assesses where progress has been made
- should support the teaching to which it relates

o Alternative Assessment

- need for assessment to be integrated with the goals of the curriculum
- learners are engaged in self-assessment

TYPES OF LANGUAGE TESTS...2

- **Proficiency tests**
 - aims to establish a test taker's readiness for a particular communicative role
 - general measure of “language ability”
 - measures a relatively stable trait
 - used to make predictions about future language performance (Hamp-Lyons, 1998)
 - high-stakes test

TYPES OF LANGUAGE TESTS...3

○ Placement Tests

- Primarily designed to measure the students' knowledge and skills in particular areas
- Provide information about the students' present level

○ Aptitude Tests

- Measure probable performance in learning FL
- Show whether the student has any specific interest in learning a language

TYPES OF LANGUAGE TESTS...4

○ Progress Tests

- Assess the level of progress which the students have made
- Used to motivate the students
- Help the teachers to assess the degree of success

TYPES OF LANGUAGE TEST (II)

Tests distinguished by linguistic levels and skills

- e.g. tests of grammar, phonology, writing, etc.

Tests distinguished by the system of scoring

- Subjective tests
- Objective tests

Tests distinguished by the nature of the test

- Discrete-point tests
- Integrative/Holistic tests
- Direct-Indirect

THE TABLE BELOW SUMMARIZES THE DIFFERENCES BETWEEN NRT & CRT

Dimension	Criterion-Referenced Tests	Norm-Referenced Tests
Purpose	To determine whether each student has achieved specific skills or concepts. To find out how much students know before instruction begins and after it has finished.	To rank each student with respect to the achievement of others in broad areas of knowledge. To discriminate between high and low achievers.
Content	Measures specific skills which make up a designated curriculum. These skills are identified by teachers and curriculum experts. Each skill is expressed as an instructional objective.	Measures broad skill areas sampled from a variety of textbooks, syllabi, and the judgments of curriculum experts.
Item Characteristics	Each skill is tested by at least four items in order to obtain an adequate sample of student performance and to minimize the effect of guessing. The items which test any given skill are parallel in difficulty.	Each skill is usually tested by less than four items. Items vary in difficulty. Items are selected that discriminate between high and low achievers.
Score Interpretation	Each individual is compared with a preset standard for acceptable achievement. The performance of other examinees is irrelevant. A student's score is usually expressed as a percentage. Student achievement is reported for individual skills.	Each individual is compared with other examinees and assigned a score--usually expressed as a percentile, a grade equivalent score, or a stanine. Student achievement is reported for broad skill areas, although some norm-referenced tests do report student achievement for individual skills.

CONSTRUCTING THE TEST

○ **Types of Test Questions:**

- **Multiple-Choice Items**
- **True-False Items**
- **Matching Items**
- **Fill-In, Completion or Short-Answer Items**
- **Essay Questions**

MULTIPLE CHOICE ITEMS

○ Advantages:

- Extremely versatile-can measure the higher level mental processes (application, analysis, synthesis and evaluation)
- A compromise between a short answer/essay and T/F item
- Can cover a wide range of content

○ Disadvantages

- Difficult to construct plausible alternative responses

TRUE-FALSE TEST ITEMS

- **Best suited for testing 3 kinds of info:**
 - Knowledge level learning
 - Understanding of misconceptions
 - When there are two logical responses
- **Advantages:**
 - Sample a large amount of learning per unit of student testing time
- **Disadvantages:**
 - Tends to be very easy and low reliability

OBJECTIVE TEST ITEM ANALYSES

- Evaluating the Effectiveness of Items..
 - Why?
 - Scientific way to improve the quality of tests and test items
 - Identify poorly written items which mislead students
 - Identify areas (competencies) of difficulty
 - Item analyses provided info. on:
 - Item difficulty
 - Item discrimination
 - Effectiveness of alternatives in MC Tests

SHORT-ANSWER ITEMS

- Two Types: (Question and Incomplete Statement)
- Advantages:
 - Easy to construct
 - Excellent format for measuring who, what, when, and where info.
 - Guessing is minimized
 - Student must know the material- rather than simply recognize the answer
- Disadvantages:
 - Grading can be time consuming
 - More than one answer can be correct

ESSAY QUESTIONS

Types of Essay Questions

○ **Extended Response Question**

- Great deal of latitude on how to respond to a question.
- Example: Discuss essay and multiple-choice type tests.

○ **Restricted Response Question**

- More specific, easier to score, improved reliability and validity
- Example: Compare and contrast the relative advantages of disadvantages of essay and multiple choice tests with respect to: reliability, validity, objectivity, & usability.

ESSAY ITEMS

○ Advantages:

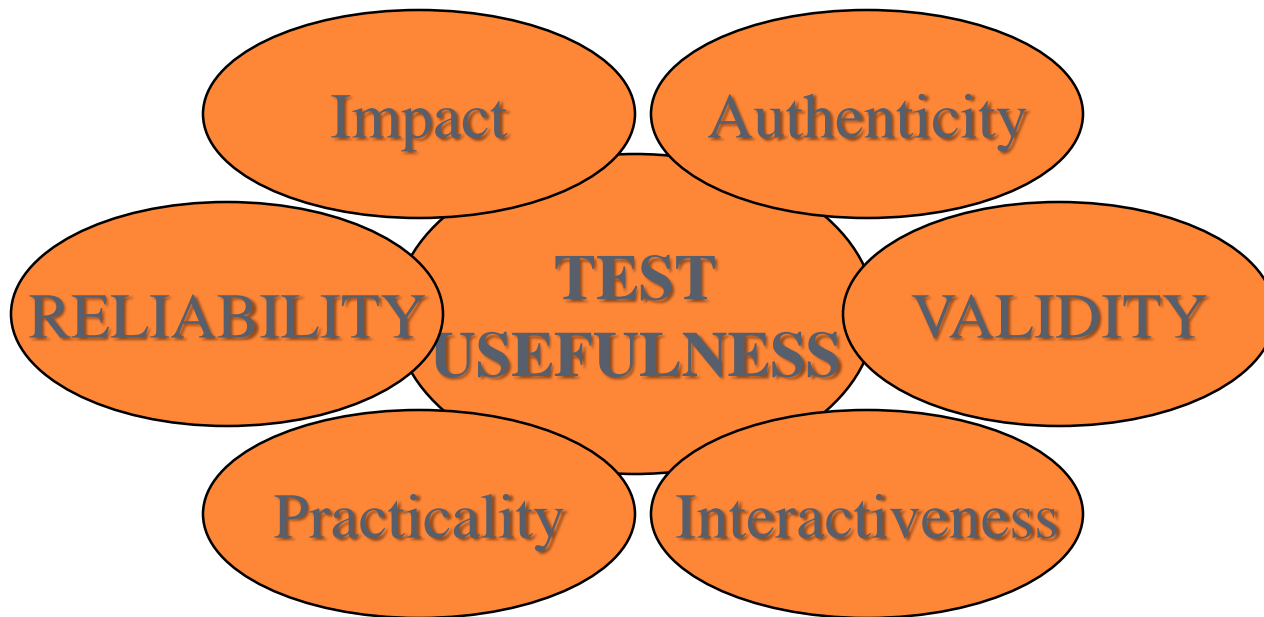
- Measures higher learning levels (synthesis, evaluation) and is easier to construct than an objective test item
- Students are less likely to answer an essay question by guessing
- Require superior study methods
- Offer students an opportunity to demonstrate their abilities to:
 - Organize knowledge
 - Express opinions
 - Foster creativity

QUESTION ARRANGEMENT ON A TEST

- **Group by question type**
 - **Common instructions will save reading time**
- **Limit the number of times students have to change frame of reference**
- **Patterns on test must be logical**
 - **Arrange from a content standpoint**
 - **Keep similar concepts together**
- **Group by difficulty (easy to hard)**

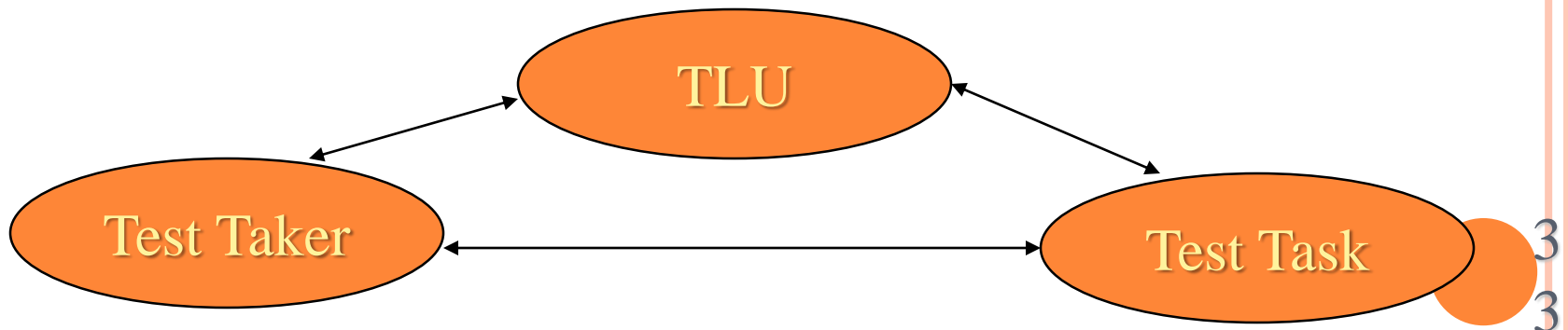
EVALUATING THE USEFULNESS OF A LANGUAGE TEST

- Usefulness= reliability+validity+ impact authenticity+interactiveness+practicality
(Bachman and Palmer, 1996)



EVALUATING THE USEFULNESS OF A LANGUAGE TEST

- Essential measurement qualities
 - reliability
 - construct validity
- Evaluation: test taker - test task - Target Language Use (TLU)



CRITICAL LANGUAGE TESTING

- Focus on consequence and ethics of test use
- Tests are embedded in cultural, educational, and political arenas whose agenda?
- Questions traditional testing knowledge
English proficiency= academic success? English: got it or get it!
- Responsible test use (Hamp-Lyons, 2000)

○ **ATTRIBUTES OF A GOOD TEST**

- **reliable: measures performance consistently**
- **valid: measures what it is supposed to and nothing else**
- **practical: constraints – time, money, expectations**
- **objective: no judgment is required from the scorer**
- **analyzable: washback**

○ **BASIC PRINCIPLES OF TEST DESIGN**

- **commercial – self-made tests**
- **item types**
- **task types and selection**
- **text types and selection**
- **dictionary use**
- **pilot**

REQUIREMENTS OF A GOOD TEST (1)

Validity: the degree to which a test measures what it is meant to measure, or can be used successfully for the purposes for which it is intended.

- Face validity
- Content validity
- Construct validity
- Empirical validity

Reliability: stability or consistency of test scores

Factors that may affect reliability include

- The extent of the sample of material selected for testing
- The administration of the test
- Scoring the test
- Test instructions
- Personal factors

REQUIREMENTS OF A GOOD TEST (2)

Discrimination: the degree to which a test or an item in a test distinguishes among better and weaker students who take the test

Practicality: the usability of a test, or practical considerations such as ease of administration, scoring and interpretation as well as financial limitations and time constraints

WASHBACK EFFECT

- “The power of tests has a strong influence on curriculum and learning outcomes”

(Shohamy, 1993)

- good test \neq positive washback
 - form of test impact depends on
 - antecedent: educational context and condition
 - process
 - consequences
- (Wall, 2000)

GOODBYE

**Any
Questions....???**

4/20/202

**Thanks for your
attention**

Tazanfal Tehseem