

Chapter One: Preliminaries of Language Testing

1. WHY TESTING

In general:

- Ascertain the degree to which goals have been realized
- Obtain precise information about candidates' abilities
- Make *educational decisions*

For students:

- *Create a positive attitude* toward class and *motivate* them in learning
- Help students prepare themselves and thus *learn the materials*

For teachers:

- Help teachers to *diagnose their efforts* in teaching
- Help teachers gain insight into ways to *improve evaluation process*

2. TEST, MEASUREMENT, EVALUATION

Measurement → the process of *quantifying* the *characteristics* of persons according to *explicit procedures and rules*.

Test → an instrument, often connoting the presentation of *a set of questions* to be answered, to obtain a measure of a characteristic of a person.

☛ **Note:** obtain *a specific sample* of behavior

Evaluation → gathering information systematically to *make decision*:

- 1) for judging decision alternatives
- 2) for determining the congruence between performance and objectives
- 3) for judging the desirability or value of a measure

3. NORM-REFERENCED TESTS vs. CRITERION-REFERENCED TESTS

Characteristics	NRT	CRT
Type of Interpretation	Relative (A student's performance is compared to those of all other students in percentile terms)	Absolute (A student's performance is compared only to the amount, or percentage, of material learned.)
Type of Measurement	To measure general language abilities or proficiencies	To measure specific objectives-based language points
Distribution of Scores	Normal distribution of scores around the mean	Varies; often non-normal

Purpose of Testing	Spread students out along a continuum of general abilities or proficiencies	Assess the amount of material known or learned by each student
Test Structure	A few relatively long subtests with a variety of item contents	A series of short, well-defined subtests with similar item contents
Knowledge of Questions	Students have little or no ideas	Students know exactly
Missed Items	It is eliminated from the test	The instructional materials are revised or additional work is given

Criterion: (1) the material that the student is supposed to learn in a course
 (2) the predetermined standard, called a criterion level

4. ASSESSMENT

Appraising or estimating the level or magnitude of some attribute of a person

4.1. Informal vs. Formal

Informal assessment → can take a number of forms, starting with unplanned comments, along with coaching and other impromptu feedback to the student, e.g., “Nice job!”; “Did you say can or can’t?”

- * **Note:** A good deal of a teacher’s informal assessment is embedded in classroom tasks
- * **Note:** Informal assessment is virtually always non-judgmental

Formal assessments → systematic, planned sampling techniques constructed to give teacher and student an appraisal of student achievement

4.2. Formative vs. Summative

Formative assessments → evaluate students in the process of ‘forming’ their competencies with the goal of helping them to continue that growth process

- * **Note:** Formative tests are either self-graded or no grade is given
- * **Note:** Virtually all kinds of informal assessment are formative

Summative assessments → given at the end of a course and the results are used primarily for assigning course grades

- * **Note:** Formative test implies the “process” of learning; summative test is concerned with the “product” of learning

5. TEACHER-MADE TESTS VS. STANDARDIZED TESTS

Characteristics	Teacher-Made Test	Standardized Test
Type of Interpretation	Criterion-referencing	Norm-referencing
Direction for Administration/Scoring	Usually no uniform directions specified	Specific, culture-free direction
Sampling of Content	Determined by classroom teacher	Determined by curriculum and subject-matter experts
Construction	May be hurried and haphazard; often no test blueprints, item tryouts, item analysis or revision	Uses meticulous construction procedures that include test blueprints, employing item tryouts, item analysis, and item revisions
Norms	Only local classroom norms	In addition to local norms, national schools district norms
Purpose and Use	Measuring particular objectives and for intra-class comparisons	Measuring broad objectives and for inter-class comparisons
Quality of Items	Unknown	High
Reliability	Unknown	High

6. THE CONSEQUENCES OF STANDARDIZED TESTING

Consequences of a test, including such considerations as its accuracy in measuring intended criteria, its impact on the preparation of test-takers, its effect on the learner, and the (intended and unintended) social consequences of a test's interpretation and use

7. WASHBACK

The effects tests have on instruction in terms of how students prepare for the test

- * **Note:** A facet of consequential validity
- * **Note:** *Cram* courses and *teaching to the test*
- * **Note:** In classroom → the information that washes back in the form of useful diagnostic information

Washback can vary in terms of *degree* (from strong to weak) and *kind* (positive or negative).

- **Harmful washback** → test content / technique is at variance with the objectives of the course
- **Beneficial washback** → a testing procedure encourages good teaching practice

8. TEST BIAS

One particular section of the candidate population is advantaged or disadvantaged by some feature of the test or item which is not relevant to what is being measured

- **Fairness** → the degree to which a test treats every student the same

9. CRITICAL TESTING

An extension of what educators call critical pedagogy

Tests are most powerful as used as the single indicators for determining the future of individuals → tests provide the mechanism for enforcing power and control

Standardized testing is the agent of cultural, social, political, educational, and ideological agendas that shape the lives of individual participants, teachers and teachers

10. AUTHENTICITY

Degree of correspondence of the characteristics of a given language test task to the features of target language task:

- The language in the test is as natural as possible.
- Items are contextualized rather than isolated.
- Topics are meaningful for the learner.
- Some thematic organization to items is provided, such as through a story line or episode.
- Tasks represent, or closely approximate, real-world tasks.