

English Language Assessment and Evaluation

BY: Gamachis Bali

Course Code: TEFL 506

Program: MA in TEFL

Credit Hours: 3

Objective & Significance

a. Objective

This course aims at equipping students with theoretical knowledge of the principles and forms of language tests. Besides, it is intended to train and give practical experience to students to construct and evaluate language tests.

b. Significance

This course is expected to be useful for providing theoretical and practical knowledge and experience to students to measure and evaluate teaching and learning achievement in a variety of forms, types, and purposes of tests based on language test construction criteria.



A Brief History on the Approaches to Language Testing

- **In the 1950s**, an era of behaviorism and special attention to contrastive analysis, testing focused on specific language elements such as the phonological, grammatical, and lexical contrasts between two languages.
- Between the **1970s** and **1980s**, communicative theories of language brought with them a more integrative view of testing in which specialists claimed that the whole of communicative event was considerably greater than the sum of its linguistic element (Clark, 1983; Brown, 2004)



Definition of Language Testing

- Test is a set of questions that have the attributes of right and wrong. Language Testing is the practice and study of evaluating the proficiency of an individual in using a particular language effectively.
- According to Oller (1979), a language testing is a device that tries to assess how much has been learned in a foreign language course, or some part of a course by learners.
- According to Brown (2004), a language testing is a method of measuring a person's ability, knowledge, or performance in a given domain.



Language Testing (Cont.)

- The activity of developing and using language tests. As a psychometric activity, language testing traditionally was more concerned with the production, development and analysis of tests.
- Recent critical and ethical approaches to language testing have placed more emphasis on the uses of language tests.
- The purpose of a language test is to determine a person's knowledge and/or ability in the language and to discriminate that person's ability from that of others.



The relationship of testing and teaching

- Both testing and teaching are so closely interrelated (or the relationship between them is as a partner).
- The test is geared to the teaching that has taken place, whereas in the latter case the teaching is often geared to the test.

“Testing and teaching are closely interrelated that it is impossible to work in either field without being constantly concerned with the other.” (Heaton, J. B. 1988)



Why test? Or Functions of testing

- Provide feedback on student learning
- Provide a basis for instructional placement
- Inform and guide instruction
- Communicate learning expectations
- Motivate and focus students' attention and effort
- Provide practice applying knowledge and skills



Why test? (Cont.)

- As devices to reinforce learning and to motivate the student
- As a mean of assessing the student's performance in the language
- As information tool for teacher to know about where the students are at the moment, to help decide what to teach next
- As information for students to find about what they know, so that they also have an awareness of what they need to learn or review.



Why test? (Cont.)

- Diagnose students' strengths and weaknesses, to identify what they know and what they don't know
- Get students to make an effort (in doing the test itself), which is likely to lead to better results and a feeling of satisfaction.
- Give students tasks which themselves may actually provide useful review or practice, as well as testing
- Provide students with a sense of achievement and progress in their learning.



Effect of testing

The effect of testing on teaching and learning is known as backwash. There are two different backwashes in the test, they are:

- The test will cause harmful backwash if the test content and testing techniques are at variance with the objective of the course.
- It will have beneficial backwash If the result of the test give an immediate effect on teaching (the syllabus was redesigned, new books were chosen, classes were conducted differently)



Testing and Evaluation

- Evaluation may focus on the effectiveness or impact of a program of instruction, examination or project.
- Students are usually not asked to evaluate while teachers carry out or take part in evaluation only in some contexts.
- ‘Experts’ or the authorities are most commonly legitimized to carry out formal evaluation.



Testing and Measurement

- Measurement is the process of determining the amount or length of something when compared with a fixed unit (e.g. using a ruler to measure length).
- It is a systematic procedure for determining the number at an object or phenomenon
- In language teaching measurement constitutes the quantification of language proficiency.
- Aspects of language knowledge, specific abilities and skills are measurable when there are transparent criteria and precise analysis of data.



Is testing synonymous with assessment ?

- No, it is not. Assessment is a more encompassing term than testing. It is the process of gathering, interpreting, and sometimes recording and using information about students' responses to an educational task in order to provide the next learning step. In other words, it is the interpretation of measurement results
- Assessment is primarily concerned with providing teachers and/or students with feedback information. In language teaching, it is a local or global procedure through which one can appraise one or more aspects of language proficiency. Assessment is transparent when clear assessment criteria have been predetermined.



Testing and assessment (Cont.)

- Is there one form of assessment? There are different forms of assessment, including: Formative assessment, Summative assessment, Self-assessment, Peer assessment.
 - The most common forms of assessment?
1. **Continuous assessment** refers to the activities required by students during the conduct of a course. It takes place within the normal teaching period and contributes to the final assessment.



Assessment (Cont.)

- 2. Formative assessment** refers to observations which allow one to determine the degree to which students know or are able to perform a given task.
- It involves all those activities (assigned by teachers and performed by students) which provide information used as feedback so that teaching may meet students' needs.
 - It can also include teacher assessment, feedback and feed-forward.
 - Focus on the processes and products of learning



Assessment (Cont.)

- Evaluation of an individual learner used to help individual improve Performance:
 - ✓ Identification of areas for improvement
 - ✓ Specific suggestions for improvement
- Develops thinking skills and helps students to be reflective learners



Assessment (Cont.)

3. Summative assessment is usually carried out at the conclusion of a unit or units of instruction, activity or plan, in order to assess acquired knowledge and skills at that particular point in time.

It usually serves the purpose of giving a grade or making a judgment about the students' achievements in the course.

Exercises or experiences specifically designed to tap into a storehouse of skills and knowledge, usually within a relatively short time limit. They are systematic, planned sampling techniques constructed to give teacher and student an appraisal of students achievement.



Assessment (Cont.)

- Evaluation of an individual learner used for judgments or decisions about the individual
 - ✓ Verification of achievement for individual
 - ✓ Motivation of individual to maintain or improve performance
 - ✓ Certification of performance
 - ✓ Grades
 - ✓ Promotion



Assessment (Cont.)

- Summative evaluation results provide both formative and summative information.
- Example: summative evaluation can be used formatively to make decisions about changes to instructional strategies, curriculum topics, or learning environment.
- Similarly, formative evaluation assists teachers in making summative judgments about student progress and determining where further instruction is necessary for individuals.
- Summative assessment is sometimes referred to as assessment of learning & formative assessment, as assessment for learning.



Summative assessment

Advantages

- Self-directed testing on various aspect of language
- Practice for high-standardized test
- Some individualization
- Large scale standardized test that can be administered easily by thousand of test taker at many different stations
- Lack of security

Disadvantages

- Occasional “home-grown” quizzes that appear on unofficial websites may be mistaken for validated assessment
- The multiple choice format proffered for most CBT contains the usual potential for flawed item design
- Open-ended response are less
- The human interactive element is absent



Assessment (Cont.)

- Computer-based testing (CBT) has emerged as one of the recent “innovative” approaches to assessments most pursued by states.
- CBT is lauded as the answer to having cheaper and speedier test delivery for state and district-wide assessments.
- It is also seen by some as an avenue toward greater accessibility for students with disabilities.
- An assessment in which test-taker performs responses on a computer

E.g. TOEFL and CAT



Other forms of assessment?

Less frequent but increasingly important forms are:

1. **Self-assessment** occurs when an appraisal instrument is self-administered for the specific purpose of providing performance feedback, diagnosis and prescription recommendations rather than a pass/fail decision.
 - Students engage in a systematic review of their progress and achievement, usually for the purpose of improvement.
 - This is similar to peer evaluation but students assess their own contribution as well as their peers using an established set of criteria.
 - It may involve comparison with an exemplar, success criteria, or other criteria. It may also involve critiquing one's own work or a description of the achievement obtained.



Other forms (Cont.)

Advantages:

- ✓ Encourages student involvement and responsibility.
- ✓ Encourages students to reflect on their role and contribution to the process of the group work.
- ✓ Allows students to see and reflect on their peers' assessment of their contribution.
- ✓ Focuses on the development of student's judgment skills.

Disadvantages:

- ✓ Potentially increases lecturer workload by needing to brief students on the process as well as on-going guidance on performing self evaluation.
- ✓ Self evaluation has a risk of being perceived as a process of presenting inflated grades and being unreliable.
- ✓ Students feel ill equipped to undertake the assessment.



Other forms (Cont.)

2. Peer assessment occurs when students judge one another's work on the basis of reference criteria. This can occur using a range of strategies.

In other words, students individually assess each other's contribution using a predetermined list of criteria. Grading is based on a predetermined process, but most commonly it is an average of the marks awarded by members of the group.

The peer assessment process needs to be taught and students need to be supported by opportunities to practice it regularly in a supportive and safe (classroom) environment.



Other forms (Cont.)

Advantages:

- Agreed marking criteria means there can be little confusion about assignment outcomes and expectations.
- Encourages student involvement and responsibility.
- Encourages students to reflect on their role and contribution to the process of the group work.
- Focuses on the development of student's judgment skills.
- Students are involved in the process and are encouraged to take part ownership of this process.
- Provides more relevant feedback to students as it is generated by their peers.
- It is considered fair by some students, because each student is judged on their own contribution.
- When operating successfully can reduce a lecturer's marking load.
- Can help reduce the 'free rider' problem as students are aware that their contribution will be graded by their peers.

Other forms (Cont.)

Disadvantages:

- Additional briefing time can increase a lecturer's workload.
- The process has a degree of risk with respect to reliability of grades as peer pressure to apply elevated grades or friendships may influence the assessment, though this can be reduced if students can submit their assessments independent of the group.
- Students will have a tendency to award everyone the same mark.
- Students feel ill equipped to undertake the assessment.
- Students may be reluctant to make judgments regarding their peers.
- At the other extreme students may be discriminated against if students 'gang up' against one group member.



Assessment Vs Testing

Assessment

- ongoing process
- generally encompasses a wide domain
- subconscious impression
- incidental judgment
- implicit evaluation

Testing

- administrative procedure
- specific times
- curriculum referenced
- peak performance
- measured and evaluated



Does assessment include testing?

- Yes, it does.
- Testing is a particular kind of assessment which focuses on eliciting a specific sample of performance.
- The implication of this is that in designing a test we construct specific tasks that will elicit performance from which we can make the inferences we want to make about the characteristics of students, groups or individuals.



Participants in Testing

- The participants in language testing are the:
 - ❖ Tester
 - ❖ Test taker/the testee
 - ❖ Test user



Tester

The tester may be:

- A foreign language teacher who designs, administers, and interprets tests given to his own learners
- A group of people responsible for developing tests requirements
- A private or governmental testing agency (PALSO in Greece; ETS – the Educational Testing Service in New Jersey, USA; CITO in Holland or BSNP in Indonesia.
- Other organizations/ international meetings: the annual Language Testing Research Colloquium, The Scientific Commission on Language Tests and Testing of the International Association of Applied Linguistics, “Language Testing” – a professional and academic journal



The Test taker/the testee

The Test Takers may be:

- Students in schools and universities
- Applicants for positions that require foreign language abilities
- People seeking certification of language proficiency for their jobs



The test user

❖ The test users are the individual or institution that make use of

- The interpretation of scores

E.g. foreign language teachers (to encourage and monitor learning, for personal feedback)

- The Ministry of Education uses tests to ensure that the National Curriculum is followed and to assess the standards achieved in school work.
- Foreign universities (American or British) use language tests (TOEFL or Cambridge Examination) to assess the proficiency and predict if applicants can attend successfully a programme of instruction in English



The test user (Cont.)

- Public and private institutions assess the linguistic competence of those employees who need a foreign language in their work.
- Foreign language teaching schools use tests for placement at an appropriate level in their courses



The beneficiaries of Testing

- Diagnostic and placement tests offer advantages of improved efficiency for learner, teacher, and educational system
- Admission tests protect admitting institutions and agencies that offer scholarships from too high a failure rate
- Certification tests offer advantages to the persons who pass the test and the agencies that hire them. They also offer protection to existing professionals organized in professional organizations who control access to certain professions
- Testing agencies TOEFL, University of Cambridge, Local Examination Syndicate, English as a Foreign Language, UCLES; tests are major sources of income for testing agencies



The overall impacts of testing on students motivation

- Testing has an impact on students' self-esteem.
- The students will be more confident of success
- Testing motivates the students to learn



How do we test?

There are different sorts of testing, including:

- Achievement testing
- Communicative testing
- Competence testing
- Diagnostic testing
- Integrative testing
- Performance testing
- Progress testing
- Proficiency testing
- Psychometric testing
- Placement Testing



Which kind of testing is the most common?

- **Achievement testing.** It is used to determine whether or not students have mastered the course content and how they should proceed. The content of achievement tests, which are commonly given at the end of the course, is generally based on the course syllabus or the course textbook.
- In other words, it is directly related to language courses, its purpose being to establish how successful individual students, group of students, or the courses themselves have been in achieving objectives (formative & summative assessment).



Achievement testing (Cont.)

- The content of the course in achievement tests are directly related to language course.
- The content of achievement tests, which are commonly given at the end of the course, is generally based on the course syllabus or the course textbook.
- **Purposes:** measure how successful individual students, groups of students, or the course themselves have been in achieving objectives.
- There are two kinds of achievement tests:
 - ✓ Final achievement test
 - ✓ Progress achievement test



Achievement testing (Cont.)

1. Final achievement test

- It's administered at the end of a course of study
- In the view of some testers, the content of a final achievement test should be based directly on a detailed course syllabus or the books and other materials used.
- The content of the test must be related to the courses with which they are concerned, but the nature of this relationship is a matter of disagreement amongst language testers.



Achievement testing (Cont.)

- Used at the end of a course
- Written by ministries of education, official examining boards, or members of teaching institutions.
- The test content can be based on a syllabus studied or a book taken during the course.
 - ✓ syllabus-content approach



Achievement testing (Cont.)

2. Progress achievement tests

- measure the progress that students are making
- based on short-term objectives



Achievement testing (Cont.)

Advantages: only contain what the learner has actually encountered = a fair test. Successful performance on the test may not truly indicate successful achievement of course objectives.

- It has their pros and cons. On the pro side, achievement tests allow for data to be collected about student achievement and learning that helps to inform educators, curriculum planners and the state officials who can help to determine the budget for a school.

Disadvantages: a badly-designed syllabus/badly-chosen book

- ✓ misleading results of a test
- ✓ unreal achievement of the course objectives



Achievement testing (Cont.)

For example:

- A course may have an objective to develop conversational abilities, but the course and the test require students only to utter statements about certain topics.
- A course may aim to develop a reading ability in German, but the test limits to the vocabulary the students have to meet.



Progress testing

- Progress testing is used at various stages throughout a language course to determine learners' progress up to that point and to see what they have learnt.
- It measure the students' improvement in relation to their syllabus. These tests only contain items which the students have been taught in class.



Progress testing

Advantages:

- ✓ provide a rich source of information
- ✓ fosters knowledge retention
- ✓ allows early detection of high achievers
- ✓ brings stability in assessment procedures
- ✓ provides excellent benchmarking opportunities

Disadvantages:

- ✓ it is very complex for reducing test difficulty in variation and standard setting procedures.
- ✓ It do not work in heterogeneous programs with early specialization.



Proficiency Testing

- Proficiency test is designed to measure people's ability in a language, regardless of any training they may have had in that language. Hughes, A. (2003)
- The content of proficiency test is NOT based on the content or objectives of language courses that people taking the test may have followed, but it is based on a specification of what candidates have to be able to do in language in order to be considered proficient.



Proficiency Testing (Cont.)

- Proficiency testing is used to measure learners' general linguistic knowledge, abilities or skills without reference to any specific course.
 - ✓ Some proficiency tests are intended to show whether students or people outside the formal educational system have reached a given level of general language ability.
 - ✓ Others are designed to show whether candidates have sufficient ability to be able to use a language in some specific area such as medicine, tourism etc. Such tests are often called Specific Purposes tests.



Proficiency Testing (Cont.)

- Other proficiency tests are designed to show whether candidates have reached a certain standard with respect to certain specified abilities.
- **Examples:**
 - ✓ Test for a United Nations translator, test for a course of study in a British university, test of arts, test of sciences.
 - ✓ The Cambridge examinations (FCE, CPE), the Oxford EFL examinations (Preliminary and Higher)



Proficiency Testing (Cont.)

Purposes:

- ✓ identify learners' strengths and weaknesses
- ✓ ascertain what learning still needs to take place

Advantages: Analysis of a learner's performance in writing and speaking in terms of grammatical accuracy or linguistic appropriacy

Disadvantages: Difficulty in identifying whether a learner masters the present perfect/past tense distinction



Which kind of testing is the least common?

- Diagnostic testing, which seeks to identify those areas in which a student needs further help.
- It's used to identify learners' strength and weaknesses.
- It's intended primarily to ascertain what learning still needs to take place.
- These tests can be fairly general, and show, for example, whether a student needs particular help with one of the four language skills; or they can be more specific, seeking to identify weaknesses in a student's use of grammar.



Diagnostic Testing

Tests for a purpose

- **Function:** measure people's ability in a language
- **Content:** base on a specification of what candidates have to be able to do to be considered proficient

E.g. test for a United Nations translator, test of arts, test of sciences

Tests for a more general purpose

- **Function:** show whether candidates have reached a certain standard
- **Content:** base on detailed specifications of *what candidates can do*

E.g. FCE, CPE



Diagnostic testing (Cont.)

- A distinct form of measurement. Its purpose is to ascertain, prior to instruction, each student's strengths, weaknesses, knowledge, and skills.
- Establishing these permits the instructor to remediate students and adjust the curriculum to meet each learner's unique needs.
- Keeping diagnostic instruments for comparison and further reference enables teachers and students to determine progress and future directions.
- Tools such as the writing strategies questionnaire and the reading interest/ attitude inventor can provide support for instructional decisions.



Psychometric testing

- Psychometric testing is aimed at measuring psychological traits such as personality, intelligence, aptitude, ability, knowledge, skills which makes specific assumptions about the nature of the ability tested.
E.g. that it is unidimensional and normally distributed.
- It includes a lot of discrete point items.



Placement test

- It's intended to provide information that will help to place students at the stage of the teaching programme most appropriate to their abilities.
- Typically it's used to assign students to classes at different level.
- Used to assign students to classes at different levels
- No one placement test will work for every institution
- The most successful one are those constructed for particular situations.



Performance Testing.

- Performance testing, which includes direct, systematic observation of an actual student performance or examples of student performances and rating of that performance according to pre-established performance criteria.
- Students are assessed on the result as well as the process engaged in a complex task or creation of a product.
- A performance test measures performance on tasks requiring the application of learning in an actual or simulated setting.
- Either the test stimulus, the desired response, or both are intended to lend a high degree of realism to the test situation.



Communicative testing

- Do all language tests aim at measuring communicative competence? No. They may test aspects of language knowledge and skills which are considered to be indicators of communicative competence.
- So, are all types of tests ‘communicative’? Tests identified as ‘communicative’ are those which are interaction-based, open-ended (that is, responses cannot be predicted as in natural communicative environments), authentic, behavior-based and so on.



Communicative testing (Cont.)

- Communicative tests are supposed to measure communicative competence which includes:
 - ✓ linguistic competence,
 - ✓ sociolinguistic competence,
 - ✓ strategic competence.



Competence Testing

- Competence testing, which is used to measure candidates' acquired capability to understand and produce a certain level of foreign language, defined by phonological, lexical grammatical, sociolinguistic and discourse constituents.
- In order to make test-takers' competence measurable and visible, testers turn of necessity to their actual performance which may indicate their competence.



Components of Communicative competence

- Grammatical competence (phonology, orthography, vocabulary, word formation, sentence formation).
- Sociolinguistic competence (social meanings, grammatical forms in different sociolinguistic contexts)
- Discourse competence (cohesion in different genres, cohesion in different genres)
- Strategic competence (grammatical difficulties, sociolinguistic difficulties, discourse difficulties, performance factors)



Characteristics of communicative tests

- **Communicative test setting requirements:**

- 1) Meaningful communication
- 2) Authentic situation
- 3) Unpredictable language input
- 4) Creative language output
- 5) All language skills

- **Bases for ratings**

- 1) Success in getting meaning across
- 2) Use focus rather than usage
- 3) New components to be rated



Language Aptitude test

- Language Aptitude test is designed to measure capacity or general ability to learn a foreign language and ultimate success in that undertaking.
- This test is strongly (ostensibly) designed to apply to the classroom learning of any language.



Language Aptitude test (Cont.)

Aptitude test can provide information that helps in determining the

- ✓ Learning readiness
- ✓ Individualizing instructions
- ✓ Organizing classroom groups
- ✓ Identifying underachievers
- ✓ Diagnosing learning problems
- ✓ Help students in their educational and vocational plans
- ✓ help to predict success to some degree
- ✓ It help to measure the probability of success in an activity



Language Aptitude test (Cont.)

Advantages

- ✓ Negates Cultural Effects
- ✓ Objective, efficient comparisons
- ✓ Standardization
- ✓ Training needs assessment
- ✓ Minimizes anxiety
- ✓ Cost effectiveness

Disadvantages

- ✓ does not necessarily result in good performance
- ✓ Strips Test-Takers of Identity
- ✓ can favor “non-thinkers
- ✓ Test Anxiety



Integrative & Discrete point testing

- Integrative testing requires a combination of many language elements in a completion of a task.
- It include activities that assess skills and knowledge in an integrated manner (e.g., reading and writing, listening and speaking).
- Less attention is paid to specific lexico-grammatical points.
- Measures several skills at one time such as dictation

E.g. writing a composition, taking a dictation or completing a cloze passage



Integrative & Discrete (Cont.)

- Discrete point testing refers to the testing of one element at a time, item by item.
- It contains items that ideally reveal the candidate's ability to handle one level of language and one element of receptive or productive skills.
- Measures the small bits and pieces of a language as in a multiple choice test made up of questions constructed to measure students' knowledge of different structures

E.g. form of series of items, each testing a particular grammar structure.

❖ *Discrete point tests* will always be *indirect* while *integrative tests* will tend to be *direct*.



Integrative & Discrete (Cont.)

Discrete

- Refers to the testing of one element at a time, item by item.
- E.g. Form of a series of items testing a particular grammatical structure
- Diagnostic tests of grammar

Integrate

- Requires the candidate to combine many language elements in the completion of a task.
- E.g. Comprehension of words and the ability to use them correctly
- Free composition
- Cloze test



Do all language tests aim at measuring communicative competence?

- No. They may test aspects of language knowledge and skills which are considered to be indicators of communicative competence.
- So, are all types of tests ‘communicative’?
- Tests identified as ‘communicative’ are those which are interaction-based, open-ended (that is, responses cannot be predicted as in natural communicative environments), authentic, behavior-based and so on.



Do all language tests aim (Cont.)

- Communicative tests are supposed to measure communicative competence which includes:
 - ✓ linguistic competence,
 - ✓ sociolinguistic competence,
 - ✓ strategic competence.



Types of Testing

Direct & Indirect Testing

- It's said to be direct when it requires the candidate to perform precisely the skill that we wish to measure.
- Direct testing is easier to carry out when it is intended to measure the productive skills of speaking and writing.
- Perform precisely the skill which is measured
- The tasks and texts should be as authentic as possible.
- Easier to measure the productive skills

E.g. To know how students pronounce a language get them to speak



Types of Testing (Cont.)

Indirect testing

- It attempts to measure the abilities underlie the skills in which we are interested.
- Measure the abilities underlying the skills

E.g. Sentence correction exercises

- ✓ an indirect measure of writing ability
- ✓ One section of TOEFL in writing section, it's developed as an indirect measure of writing ability. "at first the old woman seemed unwilling to accept anything that was offered her by my friend and I"



Advantages

Direct

- Create the conditions eliciting the behavior that the judgment is based on
- The assessment and interpretation of Ss' performance is quite straightforward
- Helpful backwash effect

Indirect

- Test a representative sample of a finite number of abilities underlying a potentially indefinite large number of manifestations of them



Disadvantages

Direct

- Small sample of tasks

Indirect

- Weak relationship between performance on test and performance of the skills



Norm referenced & Criterion Referenced Test

- Is designed to measure the global language abilities such as overall English Proficiency, academic listening ability, reading comprehension, and so on.
- Each student's score on such a test is interpreted relative to the scores of all other students who took the test with reference to normal distribution
- Criterion reference test is usually produced to measure well-defined and fairly specific instructional objectives



NRT and CRT (Cont.)

- The interpretation of CRT is considered as absolute in a sense that each student's score is meaningful without reference to the other students' scores
- The purpose of criterion–referenced testing is to *classify people according to* whether or not they are able to perform some tasks or set of tasks satisfactorily.
- Norm–referenced testing relates to one candidate's performance to that of other candidate.



NRT and CRT (Cont.)

DIMENSIONS	NRT	CRT
PURPOSE	To rank each student with respect to the achievement of others in broad areas of knowledge. - To discriminate between high and low achievers	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.
CONTENT	Measures broad skill areas sampled from a variety of textbooks, syllabi, and the judgments of curriculum experts.	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.

NRT and CRT (Cont.)

DIMENSIONS	NRT	CRT
ITEM CHARACTERISTICS	Each skill is usually tested by less than four items. - Items vary in difficulty. - Items are selected that discriminate between high and low achievers.	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.
SCORE INTERPRETATION	Student achievement is reported for broad skill areas, although some norm-referenced tests do report student achievement for individual skills.	A student's score is usually expressed as a percentage. Student achievement is reported for individual skills.



NRT and CRT (Cont.)

Characteristics	NRT	CRT
Types of interpretation	Relative	Absolute
Type of measurement	To measure general language abilities	To measure specific objective-based language points
Purpose of testing	Spread students out a long a continuum of general abilities of proficiencies	Assess the amount of material known or learned by each student
Distribution of scores	Normal distribution	Varies; often non normal.
Test structure	A few relatively long subtest with a variety of item content	A series of short-well defined subtests with similar item contents



NRT and CRT (Cont.)

NRT

- Scores are interpreted relative to each other in a *normal distribution scheme* (bell curve).
- The idea is to spread the students out on a continuum of knowledge

CRT

- Interpretation of scores is absolute and may be representational
- Measures students' ability against the predetermined standard



NRT and CRT (Cont.)

NRT

- standardized tests
- Scholastic Achievement Test
- rank order
- standard deviation
- percentile rank
- mean (average score)
- median (middle score)
- large-scale testing
- pre-determined responses

CRT

- feedback
- learning objectives
- classroom-based
- curriculum
- instructional value



Subjective Vs Objective

Subjective

- Judgment required
- If judgement is called for
- If the scorer is not looking for any one right answer

e.g. essay tests

- Different degrees of subjectivity

Objective

- No judgment required
- If no judgement is required on the part of a scorer
- Objective in the sense that there is only one answer

e.g. multiple choice tests



Subjective Vs Objective (Cont.)

- Objectivity in scoring -> reliability
- Less subjective scoring -> greater agreement (between two **different scorers; between scores of one person scoring the same test paper on different occasions**)



Communicative Language Testing

- Hymes's theory of communicative competence (1970s) influence language teaching and testing.

- ✓ Language does *not only* relate to grammar rules.
- ✓ It also involves *culturally specific rules of use*.
features of communicative context.



Communicative Language (Cont.)

Two features

1. CLTs are performance tests.

- ✓ **Assessment:** Learners are engaged in an act of communication like receptive, productive and both.

2. CLTs:

- ✓ Focus on *the social roles in real world settings*
- ✓ Offer *a means of specifying the demands of such roles in detail*

E.g. The communicative tests of English as a Foreign Language for overseas students intending to study at British universities



Communicative Language (Cont.)

- **Job analysis** is the stage in which the basis for the test design involves careful study of the communicative roles and tasks.
- Job analysis is used in the development of tests in occupational settings.

E.g. An Australian test of English as a second language for health professionals: communicating with patients, presenting cases to colleagues



What do tests do?

- What a test will appraise or measure depends on what testers wish to know and what the testers believe a test to be.
- There is indeed a difference between: Competence testing, which is used to measure candidates' acquired capability to understand and produce a certain level of foreign language, defined by phonological, lexical grammatical, sociolinguistic and discourse constituents.
- In order to make test- takers' competence measurable and visible, testers turn of necessity to their actual performance which may indicate their competence.



For whom are tests important?

- For almost all the people involved in the education process:
 - ✓ the learner who wants to know how well s/he is doing, and also wants the 'piece of paper for professional and education purposes,
 - ✓ the teacher wants to know how the learner is progressing and whether and how well s/he herself is succeeding in his job.



For whom are tests important?

(Cont.)

- ✓ the parents, who want to make sure that they're getting their money's worth,
- ✓ educational authorities and others who have some interest in the learner's progress or his/her proficiency level,
- ✓ the potential employer who relies heavily on what tests tell him/her about learner proficiency levels.



Why else is testing important?

- Because of its backwash effect. What does this mean? It is the effect that testing has on teaching. For better or worse, tests and exams exert control over what goes on in classrooms.
- This is because very many language classes are geared more or less directly to the tests or examinations the learners will end up taking. Teachers must often 'teach to' a test.



Why else is testing (Cont.)

- Is the quality of tests important for teaching? Yes.
 - ✓ If the test is a bad one (or the teacher is too narrow in his/her interpretation of it), the result may be negative wash-back, where we can say that teaching suffers because of the test coming at the end of the course.
- If the test is a good one, and its nature well understood by the teacher, the effect on the teaching may be very positive. There will be positive backwash.



Three Angels of Approach on Language Testing Research

- Language tests may be examined as tests by itself.
- It is possible to investigate learners characteristics using language tests as elicitation procedures.
- Language tests could be in specific hypotheses about psycholinguistic and sociolinguistic factors in the performance of language based task may be investigated using language tests as research tools.



Evaluation of Classroom Tests

- Are the test procedures practical?
- Is the test reliable?
- Does the procedure demonstrate content validity?
- Is the procedure face valid and biased for best?
- Are the test tasks as authentic as possible?
- Does the test give beneficial washback?



Considerations when constructing a test

- There are two basic considerations when constructing a test. It must be **valid** and **reliable**. The first concept first:
 1. **Validity**: is related to how well the tests do what it is supposed to do, namely, to inform us about the examinee's progress toward some goals in a curriculum or course study, or to differentiate levels of ability among various examinees on some tasks.



Validity (Cont.)

- In other words, validity is commonly defined as 'the extent to which [a test] measures what it is supposed to measure and nothing else.'
- If a test is valid, the outsider who looks at an individual's score knows that it is a true reflection of the individual's skill in the area the test claims to have covered.



Kinds of Validity

1. Content validity: A test is said to have content validity if the items or tasks of which it is made up constitute a representative sample of items or tasks for the area of knowledge or ability to be tested (often related to a syllabus or a course).

In other words, it requires the test taker to perform the behaviour that is being measured. Its content constitutes a representative sample of the language skills, structures, etc. With which it is meant to be measured.



Kinds of Validity (Cont.)

2. Construct validity: A test is said to have construct validity if the scores that a candidate gets on this relate in the same way to another test or form of assessment for the same aspect of knowledge.

Any theory, hypothesis, or model that attempts to explain observed phenomena in our universe and perception; proficiency and communicative competence are linguistic constructs; self-esteem and motivation are psychological constructs.



Kinds of Validity (Cont.)

In other words, a test is said to have construct validity if the scores that a candidate gets on this relate in the same way to another test or form of assessment for the same aspect of knowledge.

3. Criterion related validity: the degree to which results on the test agree with those provided by some independent and highly dependable assessment of the candidates' ability.

The extent to which the inferences made from assessment results are appropriate, meaningful and useful in terms of the purpose of the assessment.



Kinds of Validity (Cont.)

- 4. Empirical validity:** A measure of the validity of a test arrived at by comparing the test with one or more criterion measures.
- 5. Face validity:** The extent to which a test appeals to candidates or to those choosing it on behalf of the candidates because it is considered to be an acceptable measure of the ability they wish to measure. It is sometimes referred to as 'test appeal'.
- 6. Predictive validity:** A type of validity based on the degree to which a test accurately predicts future performance. A language aptitude test for example, should have predictive validity because the results of the test should predict the ability to learn a foreign language.

Kinds of Validity (Cont.)

6. Consequential validity: accuracy in measuring intended criteria, its impacts on the preparation of test takers, its effects on the learner, and social consequences of test interpretation and use.

7. Response validity [internal]: the extent to which test takers respond in the way expected by the test developers.



Kinds of Validity (Cont.)

8. Concurrent validity [external]: the extent to which test takers' scores on one test relate to those on another externally recognised test or measure.

9. Predictive validity [external]: the extent to which scores on test Y predict test takers' ability to do X

E.g. IELTS + success in academic studies at university



Kinds of Validity (Cont.)

- 'Validity is not a characteristic of a test, but a feature of the inferences made on the basis of test scores and the uses to which a test is put.'
- To make test more valid:
 - ✓ Write explicit test specification
 - ✓ Use direct testing
 - ✓ Scoring of responses related directly to what is being tested.
 - ✓ Make the test reliable.



Washback

- The quality of the relationship between a test and associated teaching.
- We have positive effect and negative effect.
- Test is valid when it has a good washback
- Students have ready access to discuss the feedback and evaluation you have given.



Washback (Cont.)

- The effect of testing on teaching and learning. The effect of test on instruction in terms of how students prepare for the test.
- Formative test: provides washback in the form of information to the learner on progress toward goals, while Summative test is always the beginning of further pursuits, more learning, more goals.
- To improve washback: use direct testing, use criterion reference-testing, base achievement tests on objectives, and make sure that the tests are understood by students and teachers.



Reliability

- Reliability is another very important consideration when testing.
- Reliability refers to the consistency of a test. That is, if every time the test is administered it will have the same outcome.
- But reliability does not have to do with the content of the test alone; it has to do with marking in two ways:
 - ✓ ensuring that different raters give comparable marks to the same script,
 - ✓ the same raters give the same marks on two different occasions to the same script.



Reliability (Cont.)

- Reliability is most often estimated with regard to:
The internal consistency in a test; that is, if there is correlation among the variables comprising the test.
- The results when testing and re-testing; that is, if there is correlation between two (or more) administrations of the same item, scale, or instrument for different times, locations, or populations, when the two administrations do not differ in other relevant variables.



Reliability (Cont.)

- Reliability: is a matter of how consistently it produces similar results on different occasions under similar circumstances. A reliable test is consistent and dependable.
- Related to accuracy, dependability and consistency.

E.g. 20°C here today, 20°C in North Italy – are they the same?

According to Henning (1987), reliability is a measure of accuracy, consistency, dependability, or fairness of scores resulting from the administration of a particular examination.

E.g. 75% on a test today, 83% tomorrow – problem with reliability.



Kinds of reliability

- 1. Inter-rater reliability:** which refers to the level of agreement between two or more evaluators/ judges/ raters on a particular instrument at a particular time.
 - They are to apply their marks in a manner that is predictable and replicable.
 - Therefore, note that inter-rater reliability is a property of the testing situation, and not of the instrument itself.



Kinds of reliability (Cont.)

- 2. Student Related reliability:** the deviation of an observed score from one's true score because of temporary illness, fatigue, anxiety, bad day, etc.
- 3. Rater reliability:** two or more scores yield an inconsistent scores of the same test because of lack attention on scoring criteria, inexperience, inattention, or preconceived bias.
- 4. Administration reliability:** unreliable results because of testing environment such as noise, poor quality of cassettee tape, etc.
- 5. Test reliability:** measurement errors because the test is too long.



To make test more Reliable

- Take enough sample of behaviour
- Exclude items which do not discriminate well between weaker and stronger students
- Do not allow candidate too much freedom.
- Provide clear and explicit instructions
- Make sure that the tests were perfectly laid out and legible
- Make candidates familiar with format and testing techniques



To make test more Reliable (Cont.)

- Provide uniform and undistracted conditions of administration
- Use items that permit objective scoring
- Provide a detailed scoring key
- Train scorers
- Identify candidate by number, not by name
- Employ multiple, independent scoring



Measuring Reliability

- **Test retest reliability:** administer whatever the test involved two times.
- **Equivalent:** forms reliability/parallel-forms reliability: administering two different but equal tests to a single group of students (e.g. Form A and B)
- **Internal consistency reliability:** estimate the consistency of a test using only information internal to a test, available in one administration of a single test. This procedure is called Split-half method.



Practicality

- Practicality: must be determined in relation to the cost in term of materials, time, and effort that it requires. It will include the preparation, administration, scoring, and interpretation of the test.
- Is not excessively expensive
- Stays within appropriate time constraints
- Is relatively easy to administer
- Has a scoring/evaluation procedure that is specific and time efficient
- Items can be replicated in terms of resources needed

Practicality (Cont.)

- Fairness issue: a test treats every student the same.
- The cost issue
- Ease of test construction
- Ease of test administration
- Ease of test scoring
- Interactions of theoretical issues

E.g. time, materials, people; can be administered, can be graded, results can be interpreted



Instructional value

- Instructional value: is pertained to how easily it can be fitted into an educational program, whether the latter involves teaching a foreign language, teaching language arts to native speakers, or verbally imparting subject matter in a monolingual or multilingual school setting.



Purposes of Language Testing

1. For an evaluation of overall proficiency.
2. To place students into a course level.
3. To measure achievement within a course.



Traditional and Alternative Assessment

1. One-shot, standardized exams
2. Time, multiple-choice format
3. Decontextualized test items
4. Scores suffice for feedback
5. Norm-referenced scores
6. Focus on the **“right” answer**
7. Summative
8. Oriented to product
9. Non-interactive performance
10. Fosters extrinsic motivation

1. Continuous long-term assessment
2. Untimed, free-response format
3. Contextualized communicative tasks
4. Individualized feedback and wash-back
5. Criterion-referenced scores
6. Open-ended, creative answers
7. Formative
8. Oriented to process
9. Interactive performance
10. Fosters intrinsic motivation



The Changing Paradigm of Assessment in the Classroom

Old Paradigm

1. Focus on language
2. Teacher-centered
3. Isolated skills
4. Emphasis on product
5. One answer, one-way correctness
6. Tests that test

New Paradigm

1. Focus on communication
2. Learner-centered
3. Integrated skills
4. Emphasis on process
5. Open-ended, multiple solutions
6. Tests that also teach



Is testing a ‘good’ or a ‘bad’ thing?

- “A well-constructed test is an instrument that provides an accurate measure of a test taker's ability within a particular domain.” Brown, 2003
- Language educators (and not only) are divided into two camps: the **teachers** and the **testers** who meet in battle.
- Teachers often say things like:
 - ✓ Let's learn to teach before we learn to test.
 - ✓ We deal with people, you deal with statistics.



Is testing a 'good' or a 'bad' (Cont.)

- Testers think that teachers:
 - ✓ tend to be unspecific about their aims and objectives.
 - ✓ are disinterested in finding out whether goals and objectives have been met.



Is testing a 'good' or a 'bad' (Cont.)

- Can we do without teaching or testing?
Probably yes because:
 - ✓ Learning can occur in spite of teaching and/or testing, despite any kind of formal evaluation.
 - ✓ The outcomes of teaching can be assessed without any form of testing.
 - ✓ Testing may be used to measure what people already know.



The elements of a good classroom test

A good test should:

- enable teachers to increase their effectiveness by making adjustments in their teaching
- help to locate the precise areas of difficulty encountered by the class or by the individual student
- enable the teacher to ascertain which parts of the language programme have been found difficult by the class
- provide the students with an opportunity to show their ability to perform a certain task



Aspects to be tested

- Four skills in communicating: listening, speaking, reading, and writing
- The language areas learnt: grammar and usage, vocabulary, and phonology
- Language elements: nouns, verbs, adjectives, and so on



Testing the language skills

- It is important to concentrate on types of test items which are relevant to the ability to use language for real-life communication, especially in oral interaction.
- Ways of assessing performance in the four major skills may take the form of tests of:
 - listening (auditory) comprehension (short utterances, dialogues, talks and lectures are given to the learners);



Testing the language skills (Cont.)

- reading comprehension (questions are set to test the students' ability to understand the gist of a text and to extract key information on specific points in the text);
- speaking ability, usually in the form of an interview, a picture description, role play and a problem-solving task involving pair work or group work;
- writing ability, usually in the form of letters, reports, memos, messages, instructions, and accounts of past events, etc.



Testing the language skills (Cont.)

- It is the test constructor's task to assess the relative importance of these skills at the various levels and to devise an accurate means of measuring the student's success in developing these skills.



Testing language areas

- In an attempt to isolate the language areas learnt, a considerable number of tests include sections on:
 - ✓ Grammar usage
 - ✓ Vocabulary (concerned with word meanings, word formation and collocations)
 - ✓ Phonology (concerned with phonemes, stress and intonation)



Testing language areas (Cont.)

- Grammar and usage: to measure students' ability to recognize appropriate grammatical forms and to manipulate structures.
- Vocabulary: to measure students' knowledge of the meaning of words and the patterns and collocations in which they occur; may test their *active or their passive vocabulary*.
- Phonology: might attempt to assess the 3 sub-skills: ability to recognize and pronounce the significant sound contrasts, ability to recognize and use the stress patterns, and ability to hear and produce the melody or patterns of the tunes (i.e. the rise and fall of the voice)



Language skills and elements

- Testing students' ability to handle the elements of the language or testing the integrated skills depends both the level and the purpose of the test.
- At all levels but the most elementary, it is generally advisable to include test items which measure the ability to communicate in the target language.



Sampling problems and avoiding traps

- The test must cover an adequate and representative section of those areas and skills which it is desired to test.
- A good test should never be constructed in such a way as to trap the students into giving an incorrect answer.



Approaches To Language Testing

1. Essay-Translation Approach
2. Structuralist Approach
3. Integrative Approach
4. Communicative Approach



Approaches To Language (Cont.)

1. Essay-Translation Approach: Characteristics and types of tests in essay-translation approach
 - This is commonly referred to as the pre-scientific stage of language testing.
 - No special skill or expertise in testing is required.
 - Tests usually consist of essay writing, translation and grammatical analysis.



Approaches To Language (Cont.)

- Tests have a heavy literary and cultural bias.
- Public examinations resulting from the tests using this approach sometimes have an oral component at the upper intermediate and advance levels.
- Pre-scientific stage of language testing;
- Subjective judgment of the teacher is considered to be of paramount importance;



Strengths of Essay-Translation Approach

- This approach is easy to follow because teachers will simply use their subjective judgement.
- The essay-translation approach may be used for testing any level of examinees.
- The model of tester can easily be modified based on the essentials of the tests.



Weaknesses of Essay-Translation Approach

- Subjective judgement of teachers tends to be biased.
- As mentioned, the tests have a heavy literary and cultural bias.



2. Structuralist Approach

Characteristics and types of tests in structuralist approach.

- This approach views that language learning is chiefly concerned with systematic acquisition of a set of habits.
- The structuralist approach involves structural linguistics which stresses the importance of constructive analysis and the need to identify and measure the learners' mastery of the separate elements of the target language such as phonology, vocabulary and grammar.



Structuralist Approach (Cont.)

- Testing the skills of listening, speaking, reading and writing is separate from another as much as possible.
- The psychometric approach to measurement with its emphasis on reliability and objectivity forms an integral part of structuralist testing.
- Chiefly concerned with the systematic acquisition of a set of habits;



Structuralist Approach (Cont.)

- Gives importance to contrastive analysis and learner's mastery of the separate elements of the target language: phonology, vocabulary and grammar;
- The skills of listening, speaking, reading and writing are tested separately or one thing at a time;
- The psychometric approach to measurement is emphasized (objective/multiple-choice test).



Strengths of Structuralist Approach

- In testing students' capability, this approach may objectively and surely be used by testers.
- Many forms of tests can be covered in the test in a short time.
- Using this approach in testing will help students find their strengths and weaknesses in every skill they study.



Weaknesses of Structuralist Approach

- It tends to be a complicated job for teachers to prepare questionnaires using this approach.
- This approach considers measuring non-integrated skills more than integrated skills.



3. Integrative Approach

Characteristics and types of tests in integrative approach

- This approach involves the testing of language in context and is thus concerned primarily with meaning and the total communicative effect of discourse.
- Integrative tests are concerned with a global view of proficiency.
- Integrative testing involves functional language but not the use of functional language.
- The use of cloze test, dictation, oral interview, translation and essay writing are included in many integrative tests.



Integrative Approach (Cont.)

- Involve the testing of language in context;
- Concerned primarily with meaning;
- Assess the learner's ability to use two or more skills simultaneously;
- Involves functional language;
- Characterized by the use of cloze testing and dictation, also oral interviews, translation and essay writing.



Strengths of Integrative Approach

- The approach to meaning and the total communicative effect of discourse will be very useful for students in testing.
- This approach can view students' proficiency with a global view.
- A model cloze test used in this approach measures the reader's ability to decode '*interrupted*' and '*mutilated*' messages by making the most acceptable substitutions from all the contextual clues available.
- Dictation, another type using this approach, was regarded solely as a means of measuring students' skills of listening comprehension.



Weakness of Integrative Approach

- Even if many think that measuring integrated skills is better, sometimes there is a need to consider the importance of measuring skills based on students' need, such as writing only, speaking only, etc.



4. Communicative Approach

Characteristics and types of tests in communicative approach

- Communicative tests are concerned primarily with how language is used in communication.
- Language use is often emphasized to the exclusion of language usage.
- The attempt to measure different language skills in communicative tests is based on a view of language referred to as divisibility hypothesis.



Communicative Approach (Cont.)

- The test content should totally be relevant for a particular group of examinees and the tasks set should relate to real-life situation.
- Communicative testing introduces the concept of qualitative modes of assessment in preference to quantitative modes of assessment.



Strengths of Communicative Approach

- Communicative tests are able to measure all integrated skills of students.
- The tests using this approach face students in real life so it will be very useful for them.



Strengths of Communicative (Cont.)

- Because a communicative test can measure all language skills, it can help students in getting the score.
- Consider students who have a poor ability in using spoken language but may score quite highly on tests of reading.
- Detailed statements of each performance level serve to increase the reliability of the scoring by enabling the examiner to make decisions according to carefully drawn-up and well-established criteria.



Weaknesses of Communicative Approach

- Unlike the structuralist approach, this approach does not emphasize learning structural grammar, yet it may be difficult to achieve communicative competence without a considerable mastery of the grammar of a language.
- It is possible for cultural bias to affect the reliability of the tests being administered.



Types of assessment tools

- When a number of assessment tools are used in conjunction with one another, richer and more in-depth data collection results.
- Whatever method of data collection is used, teachers should:
 - ✓ Meet with students regularly to discuss their progress
 - ✓ Adjust rating criteria as learners change and progress



Statements about feedback: True or false?

- The fact that the teacher gives feedback on student performance implies a power hierarchy: the teacher above, the student below.
- Assessment is potentially humiliating to the assessed person.
- Teachers should give their students only positive feedback, in order to encourage, raise confidence and promote feelings of success; negative feedback demoralizes.

Statements about feedback (Cont.)

- Giving plenty of praise and encouragement is important for the fostering of good teacher-student relationships.
- Very frequent approval and praise lose their encouraging effect; and lack of praise may then be interpreted as negative feedback.
- Teachers should not let students correct each other's work, as this is harmful to their relationships.



More than one correct answer

- The apple is located on or around
A) a table C) the table
B) an table D) table
- Two correct answers (A and C), wordy (somewhere around), repeat the word table inefficiently



Multiple Choice

- Do you see the chair and table? The apple is on _____ table.
a) A c) the
b) An d) (no article)
- Option d (no article) will be easily detected as a wrong option so it is not a good distracter.



True-False

- According to the passage, antidisestablishmentarianism diverges fundamentally from the conventional proceedings and traditions of the Church of England.
- * Containing too difficult vocabulary.



Ambiguous Word

- Why are statistical studies inaccessible to language teachers in Brazil according to the reading passage?
- Accessible: language teachers get very little training in mathematics and/or such teachers are averse to numbers
- Accessible: the libraries may be far away.



Double negatives

- One theory that is not unassociated with Noam Chomsky is:
 - A. Transformational generative grammar
 - B. Case grammar
 - C. Non-universal phonology
 - D. Acoustic phonology
- ✓ Use one negative only
- ✓ Emphasize it by underline, upper case, or bold-face.
- E.g.: not, NEVER, *inconsistent*



Receptive response items

- **True-False**

- 1) the statement worded carefully enough so it can be judged without ambiguity
- 2) absoluteness clues are avoided

- **Multiple Choice**

- 1) Unintentional clues are avoided
- 2) The distracters are plausible
- 3) Needless redundancy in the options is avoided
- 4) Ordering of the option is carefully considered
- 5) The correct answers are randomly assigned



Receptive response items (Cont.)

- **Matching**

- 1) More options than premises
- 2) Options shorter than premises to reduce reading
- 3) Option and premise lists related to one central theme



True-False

- Items should be worded carefully enough so it can be judged without ambiguity
- Avoid absoluteness
- This book is always crystal clear in all its explanation: T F
 - ✓ Allow the students to answer correctly without knowing the correct response.
 - ✓ Absolute clues: all, always, absolutely, never, rarely, most often



Multiple Choice

- Avoid unintentional clues

The fruit that Adam ate in the Bible was **an** _____

- | | |
|-----------|-----------|
| A. Pear | C. Apple |
| B. Banana | D. Papaya |

Unintentional clues: grammatical, phonological, morphological, etc.



Multiple Choice (Cont.)

Are all distracters plausible?

Adam ate _____

- A. An apple C. an apricot
B. A banana D. a tire

- Avoid needless redundancy
- The boy on his way to the store, walking down the street, when he stepped on a piece of cold wet ice and
A. fell flat on his face B. fall flat on his face
C. felled flat on his face D. falled flat on his face



Multiple Choice (Cont.)

- More effective:

The boy stepped on a piece of ice and _____
flat on his face.

A. fell B. fall C. felled D. falled

- Correct answers should be randomly assigned
- Distracters like “none of the above”, “A and B only”, “all of the above should be avoided



Matching

- Present the students with two columns of information; the students then must find and identify matches between the two sets of information.
- The information on the left-hand column is called matching-item premise
- On the right hand column is called option



Matching (Cont.)

- More options should be supplied than premises so the students can narrow down the choices as they progress through the test simply by keeping track of the options they have used.
- Options should be shorter than premises because most students will read a premise then search through the options.
- The options and premises should relate to one central theme that is obvious to students



Fill in Items

- The required response should be concise

Bad item:

- John walked down the street _____ (slowly, quickly, angrily, carefully, etc.)

Good item:

- John stepped onto the ice and immediately _____ down hard (fell)



Fill in Items (Cont.)

- There should be a sufficient context to convey the intent of the question to the students.
- The blanks should be standard in length
- The main body of the question should precede the blank
- Develop a list of acceptable responses



Short Response

- Items that the students can answer in a few phrases or sentences.
- The item should be formatted that only one relatively concise answer is possible.
- The item is framed as a clear and direct item

E.g. According to the reading passage, what are the three steps in doing research?



Task Items

- Task item is any of a group of fairly-open ended item types that require students to perform a task in the language that is being tested.
- The task should be clearly defined
- The task should be sufficiently narrow for the time available.
- A scoring procedure should be worked out in advance in regard to the approach that will be used.
- A scoring procedure should be worked out in advance in regard to the categories of language that will be rated.
- The scoring procedure should be clearly defined in terms of what each scores within each category means.
- The scoring should be anonymous



Analytic Score for Rating Composition Tasks

	20-18 Excellent to Good	17-15 Good to Adequate	14-12 Adequate to Fair	11- Unacceptabl e	5-1 Not-college level work
Organization (introduction, body, conclusion)					
Logical development of ideas					
Grammar					
Punctuation, Spelling, mechanics					
Style and quality of expressions					

Holistic Version of the Scale for Rating Composition Tasks

- Content
- Organization
- Language Use
- Vocabulary
- Mechanics



Personal Response Items

- The response allows the students to communicate in ways and about things that are interesting to them personally.
- Personal Responses include: self assessment, conferences, portfolio



Self-Assessment

- Decide on a scoring type
- Decide what aspect of students' language performance they will be assessing
- Develop a written rating for the learners
- The rating scale should decide concrete language and behaviours in simple terms
- Plan the logistics of how the students will assess themselves
- The students should the self-scoring procedures
- Have another student/teacher do the same scoring



Interviews/conferences

- Introduce and explain conferences to the students
- Give the students the sense that they are in control of the conference
- Focus the discussion on the students' views concerning the learning process
- Work with the students concerning self-image issue
- Elicit performances on specific skills that need to be reviewed.
- The conferences should be scheduled regularly



Interviews/conferences (Cont.)

- Teacher-student interviews or conferences are productive means of assessing individual achievement and needs.
- Interview questions can be developed to meet the needs of specific students and to fit the curriculum objectives.



Interviews/conferences (Cont.)

Sample questions

- Which speaking, listening, and viewing activities did you participate in this week? Which did you enjoy/ dislike? Why?
- Which oral activity did you find most difficult? Why? Did you solve the difficulties? How?
- In which speaking activity do you think you did your best? What makes you think so?
- What type of speaking activities would you like to learn to do better?



Portfolios

- Explain the portfolios to the students
 - Decide who will take responsibility for what
 - Select and collect meaningful work.
 - The students periodically reflect in writing on their portfolios
 - Have other students, teachers, outsiders periodically examined the portfolios.
-
- Collections of relevant works that reflect student's individual efforts, development, and progress over a designated period of time
 - Provide students, teachers, parents, and administrators with a broad picture of each student's growth



Observation

- Occurs during the student's daily practice of the four macro skills.
- Can be recorded as anecdotal notes and on checklists or rating scales.



Anecdotal Records

- Notes written regarding student language behavior, or learning.
- Document and describe significant daily events, and relevant aspects of student activity and progress.



Checklists

- Usually completed while students are engaged in specific activities or processes.
- Lists of specific criteria that teachers focus on at a particular time or during a particular process.
- Inform teachers about where their instruction has been successful and where students need assistance or further instruction



Rating scales

- Rating scales record the extent to which the criteria have been achieved by the student or are present in the student's work.
- It records the quality of the student's performance at a given time or within a given process.



Rubrics

- Rubrics include criteria that describe each level of the rating scale and are used to determine student progress in comparison to the expectations.



General Guidelines for Item Formats

- Correctly matched to the purpose and content of the item
- Only one correct answer?
- Written at the students' level of proficiency
- Avoiding ambiguous terms and statements
- Avoiding negatives and double negatives
- Avoid giving clues that could be used in answering other items
- All parts of the item on the same page
- Only relevant information presented
- Avoiding bias of race, gender and nationality
- Let another person look over the item



Term Paper #1

- How does one define what will be tested?
- How can tests and feedback provided be a positive asset in the educational process?
- What types of feedback can teachers provide to test-takers and how?
- Are tests the best tools for evaluation and assessment?
- What are the most important things to remember about raters and marking (closed and open-ended items) in integrative and discrete point tests?



Term Paper #2

- Based on the general guidelines for item formats, please prepare all types of exam questions that host both macro and micro skills from grades 5 to 12 according to the distribution I have made below.

1. Abraham Kilta = 5
2. Desta Wako = 8
3. Getachew Elias = 7
4. Mezid Endris = 6
5. Shendu Haro = 12
6. Tilahun Kassahun = 11
7. Ture Hirbaye = 10
8. Yibeltal = 9



References

- Alderson, J. C., & Banerjee, J. (2001). Language testing and assessment (Part I). *Language teaching*, 34(4), 213-236.
- Alderson, J. C., & Banerjee, J. (2002). Language testing and assessment (Part 2). *Language teaching*, 35(2), 79-113.
- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in Practice: design and developing useful language tests*. Oxford: Oxford University Press.
- Bachman, L. F. (1997). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. New York: Routledge.
- Heaton, J. B. (1988). *Writing English language tests (New ed.)*. London: Longman.



THE END!

Thank You for Attention

