Language Testing and Evaluation

Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. New York: Pearson Education.

Prof. Dr. Sabri KOÇ

1 Testing, Assessing, and Teaching

WHAT IS A TEST?

A test, in simple terms, is a *method of measuring a person's ability, knowledge, or performance in a given domain.*

A test is first a **method**. It is an instrument - a set of techniques, procedures, or items - that requires performance on the part of the test-taker. **To qualify as a test, the method must be explicit and structured:**

- multiple-choice questions with prescribed correct answers;
- a writing prompt with a scoring rubric;
- an oral interview based on a question script and a checklist of expected responses to be filled in by the administrator.

1 Testing, Assessing, and Teaching

WHAT IS A TEST?

Second, a test must **measure**.

- Some tests measure general ability, while others focus on very specific competencies or objectives.

- A multi-skill proficiency test determines general ability.

- A quiz on recognizing correct use of definite articles measures specific knowledge.

- If an instrument does not specify a form reporting measurement, then that technique cannot be defined as a test.

1 Testing, Assessing, and Teaching

WHAT IS A TEST?

- Next, a test measures **an individual's ability, knowledge or performance.** Testers need to understand who the test-takers are. What is their previous experience and background? Is the test appropriately matched to their abilities? How should test-takers interpret their scores?

- A test measures **performance**, but the results imply the test-taker's ability, or competence. Most language tests measure one's ability to perform language, that is, to speak, write, read, or listen to a subset of language.

1 Testing, Assessing, and Teaching

WHAT IS A TEST?

- Finally, a test measures a given **domain**. A proficiency test the domain is overall proficiency in a language - general competence in all skills of a language. Other tests may have more specific criteria. A test of pronunciation may include only a limited set of phonemic minimal pairs. A vocabulary test may focus only the set of words covered in a particular lesson or unit.

1 Testing, Assessing, and Teaching

WHAT IS A TEST?

- **A well-constructed test is an instrument that provides an accurate measure of the test-taker's ability within a particular domain.** The definition sounds fairly simple, but in fact, constructing a good test is a complex task involving both science and art.

1 Testing, Assessing, and Teaching

ASSESSMENT AND TEACHING

- **Assessment** is a popular and sometimes misunderstood term in current educational practice. Testing and assessing are not synonymous terms.

- **Tests** are administrative procedures that occur at identifiable times in a curriculum when learners put together all their faculties to offer peak performance, knowing that their responses are being measured and evaluated.
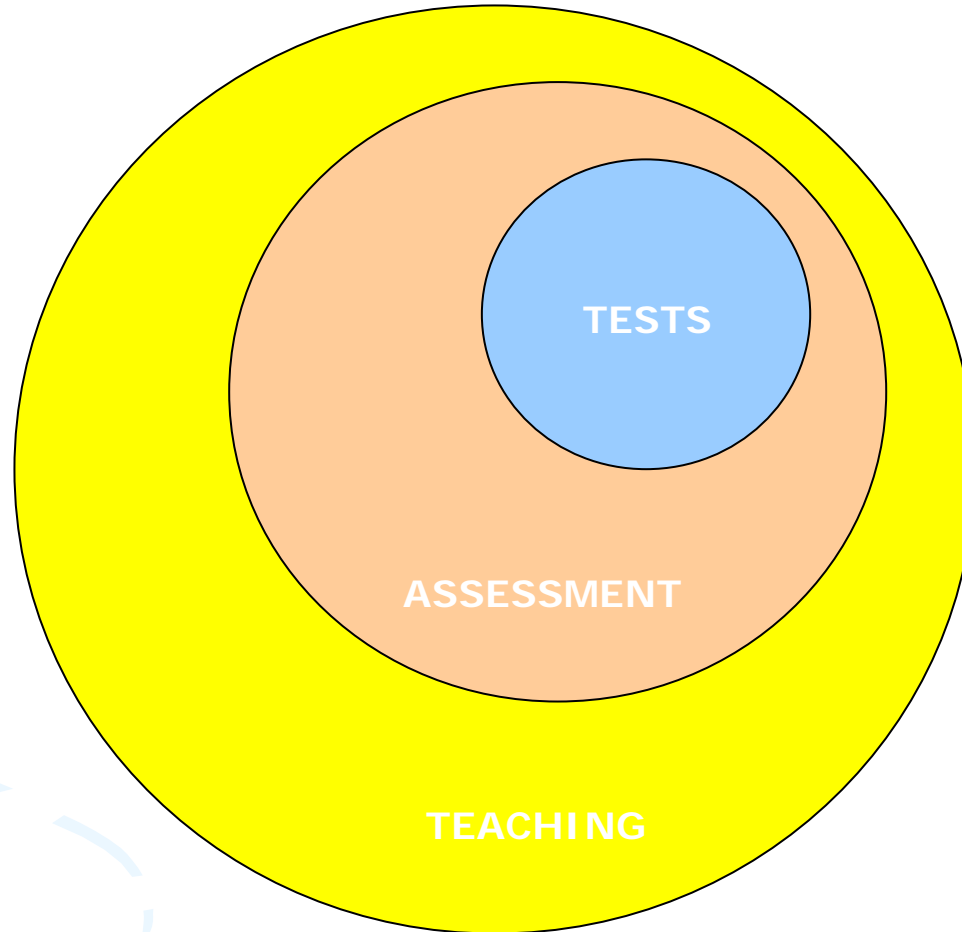
1 Testing, Assessing, and Teaching

ASSESSMENT AND TEACHING

- **Assessment**, on the other hand, is an ongoing process that encompasses a much wider domain.

- Any kind of evaluation or judgment of the student's performance is assessment.

- **Tests are a subset of assessment.** They **are not** the only form of assessment. (See the diagram on the next slide)

1 Testing, Assessing, and Teaching

Figure 1.1 Tests, assessment, and teaching (H. D. Brown, 2004:5)

1 Testing, Assessing, and Teaching

ASSESSMENT AND TEACHING

- For optimal learning to take place, students in the classroom must have the freedom to experiment, to try out their hypotheses about language without feeling that their overall competence is being judged in terms of those trials and errors. They must have ample opportunities to "play" with language in a classroom without being formally graded. **Teaching** sets up the practice games of language learning: the opportunities for learners to listen, think, take risks, set goals, and process feedback from the teacher.

1 Testing, Assessing, and Teaching

Informal and Formal Assessment

- **Informal assessment** can take a number of forms, starting with incidental, unplanned comments and responses, along with coaching and other impromptu feedback to the student. A good deal of a teacher's informal assessment is embedded in classroom tasks designed to elicit performance without recording results and making fixed judgments about the student's competence.

- **Formal assessments** are exercises or procedures specifically designed to tap into a storehouse of skills and knowledge. They are systematic, planned sampling techniques constructed to give teacher and student an appraisal of student achievement.

1 Testing, Assessing, and Teaching

Informal and Formal Assessment

What is the relationship between test and formal assessment?

- All **tests** are **formal assessments**, but not all formal assessment is testing. There are other forms of assessment such as the evaluation of journals or portfolios, a systematic set of observations of a student's frequency of oral participation in class.

1 Testing, Assessing, and Teaching

Formative and Summative Assessment

- Most of our classroom assessment is **formative assessment**: evaluating students in the process of "forming" their competences and skills with the goal of helping them continue that growth process. For all practical purposes, virtually all kinds of informal assessment are **formative**.

- **Summative assessment** aims to measure, or summarize, what a student has grasped, and typically occurs at the end of a course or unit of instruction. A summation of what a student has learned implies looking back and taking stock of how well that student has accomplished objectives, but does not necessarily point the way to future progress. Final exams in a course and general proficiency exams are examples of summative assessment.

1 Testing, Assessing, and Teaching

Norm-Referenced and Criterion-Referenced Tests

In norm-referenced tests, each test-taker's score is interpreted in relation to a mean (average score), median (middle score), standard deviation (extent of variance in scores), and/or percentage rank. The purpose in such tests (SAT, TOEFL) is to place test-takers along a mathematical continuum in rank order (numerically).

1 Testing, Assessing, and Teaching

Norm-Referenced and Criterion-Referenced Tests

Criterion-referenced tests are designed to give test-takers feedback, usually in the form of grades, on specific course or lesson objectives. Classroom tests involving the students in only one class, and connected to a curriculum, are typical of criterion-referenced testing. This type is more prominent.

1 Testing, Assessing, and Teaching

APPROACHES TO LANGUAGE TESTING: A BRIEF HISTORY

- **The 50s: Behaviorism:** Attention to contrastive analysis: focus on specific language elements as focus on specific elements such as phonological, grammatical and lexical contrasts between two languages.
- **The 70s and 80s:** Communicative theories of language: testing communicative events.
- **Today:** Communicative but with more emphasis on more authentic, valid instruments that stimulate real-world interaction.

1 Testing, Assessing, and Teaching

Discrete-point and Integrative Testing

- **Discrete-point tests** are constructed on the assumption that language can be broken down into its component parts (skills or language content) and those parts can be tested successfully.
- **Problem: decontextualization**

- **Integrative testing.** Claim: communicative competence is so global and requires such integration that it cannot be captured in additive tests of language content or other discrete points of language.
    - **A cloze test**
    - **Dictation**

1 Testing, Assessing, and Teaching

- **Unitary trait hypothesis** suggested an "indivisible" view of language proficiency: that the language content or the four skills and other discrete points of language could not be disintegrated from each other in language performance.

- **Communicative language testing**
- The mid 80s: unitary trait hypothesis was abandoned and people began designing communicative language testing.

1 Testing, Assessing, and Teaching

Need for a correspondence between language test performance and language use: Communicative language testing.

- **Test designers centered on communicative performance. Language competence consisting of organizational and pragmatic competence, later strategic competence (the ability to employ communicative strategies to compensate for breakdowns as well as to enhance the rhetorical effect of utterances) in the process of communication.**

- **Test constructors began to identify the kinds of real-world tasks that language learners were called upon to perform.**

1 Testing, Assessing, and Teaching

Performance-based assessment

Instead of just offering paper-and-pencil selective response tests of many separate items, performance-based assessment of language typically involves oral production, written production, open-ended responses, integrated performance (across skill areas), group performance, and other integrative tasks.

1 Testing, Assessing, and Teaching

Performance-based assessment

- A characteristic of many performance-based language assessments: **presence of interactive tasks (speaking, requesting, responding, or in combining listening and speaking, reading and writing.** (oral interview)

- **Performance-based assessment** means that the tester will find it difficult to distinguish between formal and informal assessment.

1 Testing, Assessing, and Teaching

CURRENT ISSUES IN CLASSROOM TESTING

- **Both communicative and performance-based assessments continue to challenge.**

- **Three issues will be discussed here:**

- **The effect of new views of intelligence on the testing industry**

- **Traditional and alternative assessment**

- **Computer-based testing**

1 Testing, Assessing, and Teaching

CURRENT ISSUES IN CLASSROOM TESTING

New Views on Intelligence

Intelligence was once considered as the ability to perform:
(a) linguistic and
(b) logical-mathematical problem solving.

Umbrella term: IQ (intelligence quotient)

1 Testing, Assessing, and Teaching

CURRENT ISSUES IN CLASSROOM TESTING

Gardner (1983, 1999) presented his **Multiple Intelligences Theory** which turned the psychometric world upside down.

Although Gardner accepted the traditional conceptualizations of linguistic and logical mathematical intelligence on which IQ tests are based, he included other "frames of mind" in his theory of multiple intelligences:

1 Testing, Assessing, and Teaching

CURRENT ISSUES IN CLASSROOM TESTING

- **spatial intelligence** (the ability to find your way around an environment, to form mental images of reality)
- **musical intelligence** (the ability to perceive and create pitch and rhythmic patterns)
- **bodily-kinesthetic intelligence** (fine motor movement, athletic prowess)
- **interpersonal intelligence** (the ability to understand others and how they feel, and to interact effectively with them)
- **intrapersonal intelligence** (the ability to understand oneself and to develop a sense of self-identity)

- **What other intelligences do you know?**

1 Testing, Assessing, and Teaching

Eight Ways of Learning

Children who are highly . . .	Think . . .	Love . . .	Need . . .
Linguistic	in words	reading, writing, telling stories, playing word games	books, tapes, writing tools, paper, diaries, dialogue, discussion, debate, stories
Logical-Mathematical	by reasoning	experimenting, questioning, figuring out logical puzzles, calculating	materials to experiment with, science materials, manipulatives, trips to planetariums and science museums
Spatial	in images and pictures	designing, drawing, visualizing, doodling	art, Legos, videos, movies, slides, imagination games, mazes, puzzles, illustrated books, trips to art museums
Bodily-Kinesthetic	through somatic sensations	dancing, running, jumping, building, touching, gesturing	role-play, drama, movement, building things, sports and physical games, tactile experiences, hands-on learning
Musical	via rhythms and melodies	singing, whistling, humming, tapping feet and hands, listening	sing-along time, trips to concerts, playing music at home and school, musical instruments
Interpersonal	by bouncing ideas off other people	leading, organizing, relating, manipulating, mediating, partying	friends, group games, social gatherings, community events, clubs, mentors/apprenticeships
Intrapersonal	in relation to their needs, feelings, and goals	setting goals, meditating, dreaming, planning, reflecting	secret places, time alone, self-paced projects, choices
Naturalist	through nature and natural forms	playing with pets, gardening, investigating nature, raising animals, caring for planet earth	access to nature, opportunities for interacting with animals, tools for investigating nature (e.g., magnifying glasses, binoculars)

How do you understand which intelligences you have?

Click here to do the MI Inventory.

1 Testing, Assessing, and Teaching

CURRENT ISSUES IN CLASSROOM TESTING
- Sternberg (1988, 1997) also charted new territory in intelligence research in recognizing creative thinking and manipulative strategies as part of intelligence. All 'smart' people are not necessarily adept at fast, reactive thinking. They may be very innovative in being able to think beyond the normal limits imposed by existing tests, but they may need a good deal of processing time to enact this creativity.

- Goleman's (1995) **concept of EQ (Emotional quotient):** We underscore the importance of the emotions in our cognitive processing. Those who manage their emotions tend to be more capable of fully intelligent processing. So control your anger, grief, resentment, self-doubt and other emotions and achieve high performance in everyday tasks.

1 Testing, Assessing, and Teaching

CURRENT ISSUES IN CLASSROOM TESTING

Traditional and "alternative" assessment

- **Trend: to supplement traditional test designs with alternatives that are more authentic in their elicitation of meaningful communication.**

- **See the next page for differences between traditional and alternative assessment.**

1 Testing, Assessing, and Teaching

Table 1.1. Traditional and alternative assessment

Traditional Assessment	Alternative Assessment
One-shot, standardized exams	Continuous long-term assessment
Timed, multiple-choice format	Untimed, free-response format
Decontextualized test items	Contextualized communicative tasks
Scores suffice for feedback	Individualized feedback and washback
Norm-referenced scores	Criterion-referenced scores
Focus on the "right" answer	Open-ended, creative answers
Summative	Formative
Oriented to product	Oriented to process
Non-interactive performance	Interactive performance
Fosters extrinsic motivation	Fosters intrinsic motivation

1 Testing, Assessing, and Teaching

CURRENT ISSUES IN CLASSROOM TESTING

Computer-Based Testing

- In a **computer-adaptive test (CAT)**, each test-taker receives a set of questions that meet the test specifications and that are generally appropriate for his or her performance level. It starts with questions of moderate difficulty. The computer is programmed to fulfill the test design as it continuously adjusts to find questions of appropriate difficulty for test-takers at all performance levels.

1 Testing, Assessing, and Teaching

Advantages of Computer-Based Testing (CAT):

- classroom-based testing

- self-directed testing on various aspects of language

- Practice for upcoming standardized tests

- some individualization

- large-scale standardized tests can be administered to thousands of test-takers and scored electronically.

1 Testing, Assessing, and Teaching

CURRENT ISSUES IN CLASSROOM TESTING

Disadvantages of Computer-Based Testing (CAT):

- lack of security and cheating

- possibility of low quality test websites

- problems with multiple-choice tech.

- subjectivity with open-ended items

- lack of human interaction

1 Testing, Assessing, and Teaching

CURRENT ISSUES IN CLASSROOM TESTING

Basic principles of effective testing:

- assessments can increase motivation

- they aid in reinforcement and retention of information

- they diagnose weaknesses and strengths

- they can provide a sense of periodic closure to modules within curriculum

- they can promote student autonomy through self-evaluation

- they can motivate learners to set goals for themselves

- they can aid in evaluating teaching effectiveness

THANK YOU FOR YOUR PATIENCE!

An MI Inventory for Adults

Check those statements that apply in each intelligence category. Space has been provided at the end of each intelligence for you to write additional information not specifically referred to in the inventory items.

1. Linguistic Intelligence

____ Books are very important to me.

____ I can hear words in my head before I read, speak, or write them down.

____ I get more out of listening to the radio or a spoken-word recording than I do from television or films.

____ I enjoy word games like Scrabble, Anagrams, or Password.

____ I enjoy entertaining myself or others with tongue twisters, nonsense rhymes, or puns.

____ Other people sometimes have to stop and ask me to explain the meaning of the words I use in my writing and speaking.

____ English, social studies, and history were easier for me in school than math and science.

____ Learning to speak or read another language (e.g., French, Spanish, German) has been relatively easy for me.

____ My conversation includes frequent references to things that I've read or heard.

____ I've written something recently that I was particularly proud of or that earned me recognition from others.

Other Linguistic Abilities: _____

2. Logical-Mathematical Intelligence

____ I can easily compute numbers in my head.

____ Math and/or science were among my favorite subjects in school.

____ I enjoy playing games or solving brainteasers that require logical thinking.

____ I like to set up little "what if" experiments (for example, "What if I double the amount of water I give to my rosebush each week?")

____ My mind searches for patterns, regularities, or logical sequences in things.

____ I'm interested in new developments in science.

____ I believe that almost everything has a rational explanation.

____ I sometimes think in clear, abstract, wordless, imageless concepts.

____ I like finding logical flaws in things that people say and do at home and work.

____ I feel more comfortable when something has been measured, categorized, analyzed, or quantified in some way.

Other Logical-Mathematical Abilities: _____

3. Spatial Intelligence

___ I often see clear visual images when I close my eyes.

___ I'm sensitive to color.

___ I frequently use a camera or camcorder to record what I see around me.

___ I enjoy doing jigsaw puzzles, mazes, and other visual puzzles.

___ I have vivid dreams at night.

___ I can generally find my way around unfamiliar territory.

___ I like to draw or doodle.

___ Geometry was easier for me than algebra in school.

___ I can comfortably imagine how something might appear if it were looked

down on from directly above in a bird's-eye view.

___ I prefer looking at reading material that is heavily illustrated.

Other Spatial Abilities:_____

4. Bodily-Kinesthetic Intelligence

____ I engage in at least one sport or physical activity on a regular basis.

____ I find it difficult to sit still for long periods of time.

____ I like working with my hands at concrete activities such as sewing, weaving, carving, carpentry, or model building.

____ My best ideas often come to me when I'm out for a long walk or a jog or when I'm engaging in some other kind of physical activity.

____ I often like to spend my free time outdoors.

____ I frequently use hand gestures or other forms of body language when conversing with someone.

____ I need to touch things in order to learn more about them.

____ I enjoy daredevil amusement rides or similar thrilling physical experiences.

____ I would describe myself as well coordinated.

____ I need to practice a new skill rather than simply reading about it or seeing a video that describes it.

Other Bodily-Kinesthetic Abilities: _____

5. Musical Intelligence

____ I have a pleasant singing voice.

____ I can tell when a musical note is off-key.

____ I frequently listen to music on radio, records, cassettes, or compact discs.

____ I play a musical instrument.

____ My life would be poorer if there were no music in it.

____ I sometimes catch myself walking down the street with a television jingle or other tune running through my mind.

____ I can easily keep time to a piece of music with a simple percussion instrument.

____ I know the tunes to many different songs or musical pieces.

____ If I hear a musical selection once or twice, I am usually able to sing it back fairly accurately.

____ I often make tapping sounds or sing little melodies while working, studying, or learning something new.

Other Musical Abilities: _____

6. Interpersonal Intelligence

____ I'm the sort of person that people come to for advice and counsel at work or in my neighborhood.

____ I prefer group sports like badminton, volleyball, or softball to solo sports such as swimming and jogging.

____ When I have a problem, I'm more likely to seek out another person for help than attempt to work it out on my own.

____ I have at least three close friends.

____ I favor social pastimes such as Monopoly or bridge over individual recreations such as video games and solitaire.

____ I enjoy the challenge of teaching another person, or groups of people, what I know how to do.

____ I consider myself a leader (or others have called me that).

____ I feel comfortable in the midst of a crowd.

____ I like to get involved in social activities connected with my work, church, or community.

____ I would rather spend my evenings at a lively party than stay at home alone.

Other Interpersonal Abilities: _____

7. Intrapersonal Intelligence

____ I regularly spend time alone meditating, reflecting, or thinking about important life questions.

____ I have attended counseling sessions or personal growth seminars to learn more about myself.

____ I am able to respond to setbacks with resilience.

____ I have a special hobby or interest that I keep pretty much to myself.

____ I have some important goals for my life that I think about on a regular basis.

____ I have a realistic view of my strengths and weaknesses (borne out by feedback from other sources).

____ I would prefer to spend a weekend alone in a cabin in the woods rather than at a fancy resort with lots of people around.

____ I consider myself to be strong willed or independent minded.

____ I keep a personal diary or journal to record the events of my inner life.

____ I am self-employed or have at least thought seriously about starting my own business.

Other Intrapersonal Abilities: _____

8. Naturalist Intelligence

____ I like to spend time backpacking, hiking, or just walking in nature.

____ I belong to some kind of volunteer organization related to nature (e.g., Sierra Club),

and I'm concerned about helping to save nature from further destruction.

____ I thrive on having animals around the house.

____ I'm involved in a hobby that involves nature in some way (e.g., bird watching).

____ I've enrolled in courses relating to nature at community centers or colleges (e.g.,

botany, zoology).

____ I'm quite good at telling the difference between different kinds of trees, dogs, birds,

or other types of flora or fauna.

____ I like to read books and magazines or watch television shows or movies that feature

nature in some way.

____ When on vacation, I prefer to go off to a natural setting (park, campground, hiking

trail) rather than to a hotel/resort or city/cultural location.

____ I love to visit zoos, aquariums, or other places where the natural world is studied.

____ I have a garden and enjoy working regularly in it.

Other Naturalist Abilities: _____

The following intelligences are in process of development:

  9. Existentialist Intelligence
10. Culinary Intelligence
11. Mechanical Intelligence