

# Cornerstones of assessment

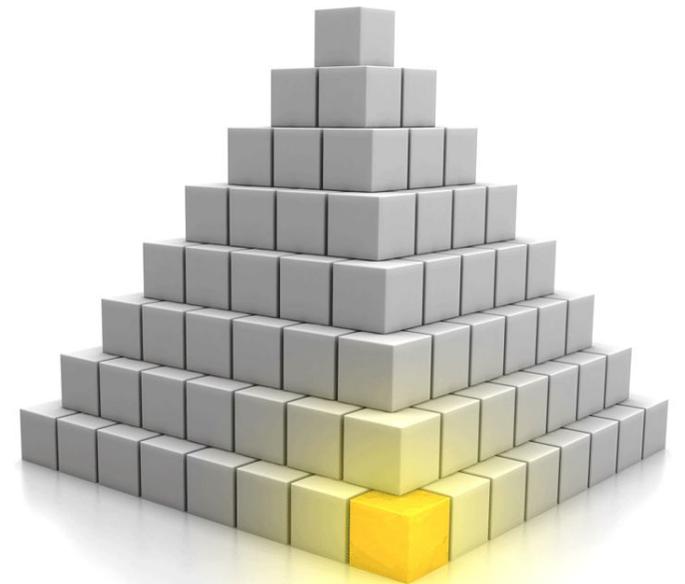
Session 2 of 11  
Assessment and International  
Exams in TEFL

## Lecture outline:

- ▶ have a basic understanding of the key principles of testing
- ▶ know why these principles are important for creating a test that is fit for purpose
- ▶ be able to assess a test according to these basic principles

# Cornerstones of Assessment

- ▶ Assessment and testing: many forms, same principles
- ▶ A good test is useful, i.e.
  - ▶ Valid and reliable
  - ▶ Practical
  - ▶ Impactful
  - ▶ Fair and secure
  - ▶ Authentic



# 1. Validity

- ✓ **Validity** - a degree to which the test actually measures what it is intended to measure.
- ✓ Test scores reflect the achievement of learning outcomes and test-taker's ability.
- ✓ The test is **valid** when it reflects what the learners can do in a language.

# Construct

- ▶ A test construct is *a latent trait*, an inherent or unobservable ability a test is trying to measure.
- ▶ Examples of constructs: math, intelligence, personality, anxiety, reading ability, pronunciation.
- ▶ Construct validity - does a test really assess the test construct?

# Construct Validity

- ▶ Grammar and Vocabulary - an essay or multiple-choice?
- ▶ Reading - reading aloud or texts and comprehension questions?
- ▶ Listening - a lecture or a series of dialogues?
- ▶ Writing ability - a dictation or a cover letter?
- ▶ Speaking - reading aloud tasks or face-to-face interviews?

# Content validity

- ▶ Assessment of course content with clear reference to goals and outcomes
- ▶ Use of formats and tasks familiar to students

# Face validity

- ▶ The test *looks* as if it measures what it is supposed to measure.
- ▶ A test must assess linguistic ability, or it may not be accepted by test-takers
- ▶ A test must look formal
  - ▶ Avoid hand-written instructions
  - ▶ Carefully introduce and explain novel assessment procedures

# To sum up on validity:

- ▶ Does the test assess the skill (construct) that you focus on in your class?
- ▶ Does the test cover the content that you have been teaching?
- ▶ Does the test look as if it is testing what it is supposed to be testing?
- ▶ It is challenging / formal / adequate enough in the eyes of the test-takers?

## 2. Reliability

- ▶ Sources of unreliability
  - ▶ Test reliability
  - ▶ Administration of test reliability
  - ▶ Consistency of results / scorer reliability
  - ▶ *Fluctuations in the learner*

# Test reliability - 1. Extent of sample material

Each new test item - **a fresh start** for the test taker

- On a reading test: “Where did the thief hide the jewels?”, “What was unusual about the hiding place?”
- + On a writing or oral production test: the more passages the test taker has to produce, the more reliable the test result is

## Test reliability - 2. Extent of freedom

1. Write a composition on tourism.
2. Write a composition on tourism in your region.
3. Write a composition on how we can develop tourism in your region.
4. Discuss the following measures intended to increase the number of foreign tourists in your region: a) better advertising and information (where? What form should it take?) b) improve facilities (hotels, transportation etc) c) training of personnel (guides, hotel managers).

## Test reliability - 3. Clear instructions

### Paraphrase using one word:

- ▶ What are you going to do after you finish university?
- ▶ Business ethics is a very difficult subject.
- ▶ You do not need to get a student ID card to access the university library.
- ▶ When I started college, the pay was \$350 a quarter.

## 4. Test administration reliability

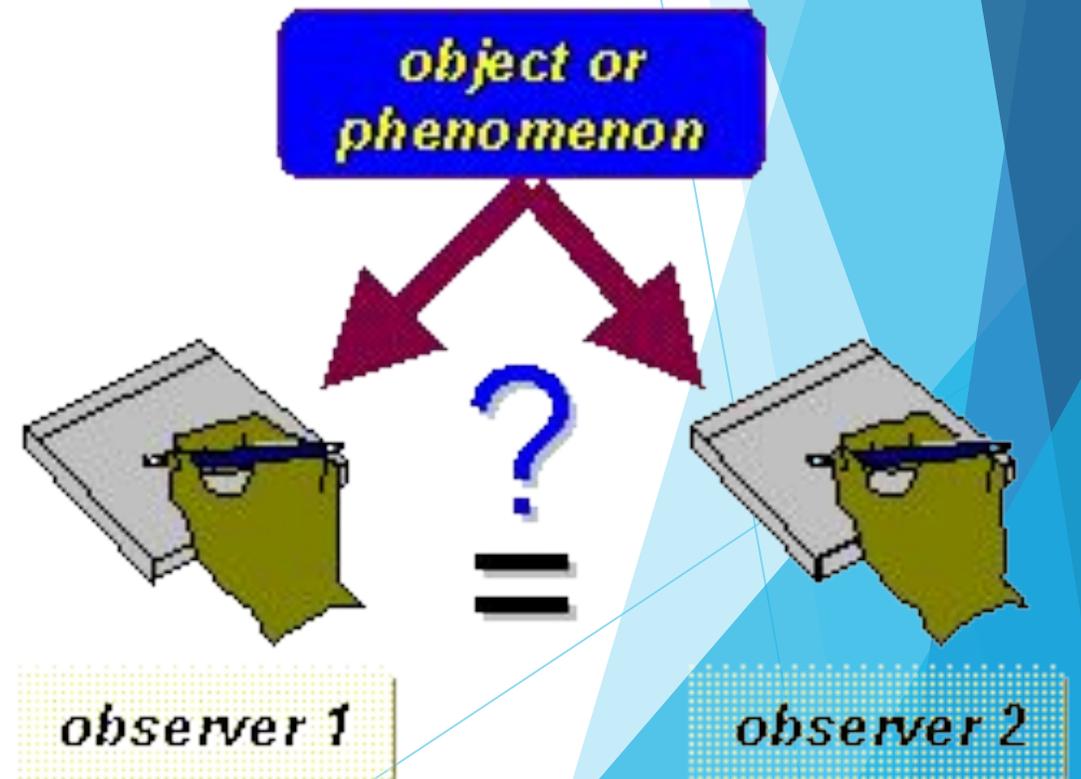
1. Layout and legibility
2. Test format and techniques
3. Uniform conditions for all test-takers



# Scorer / Inter-rater reliability

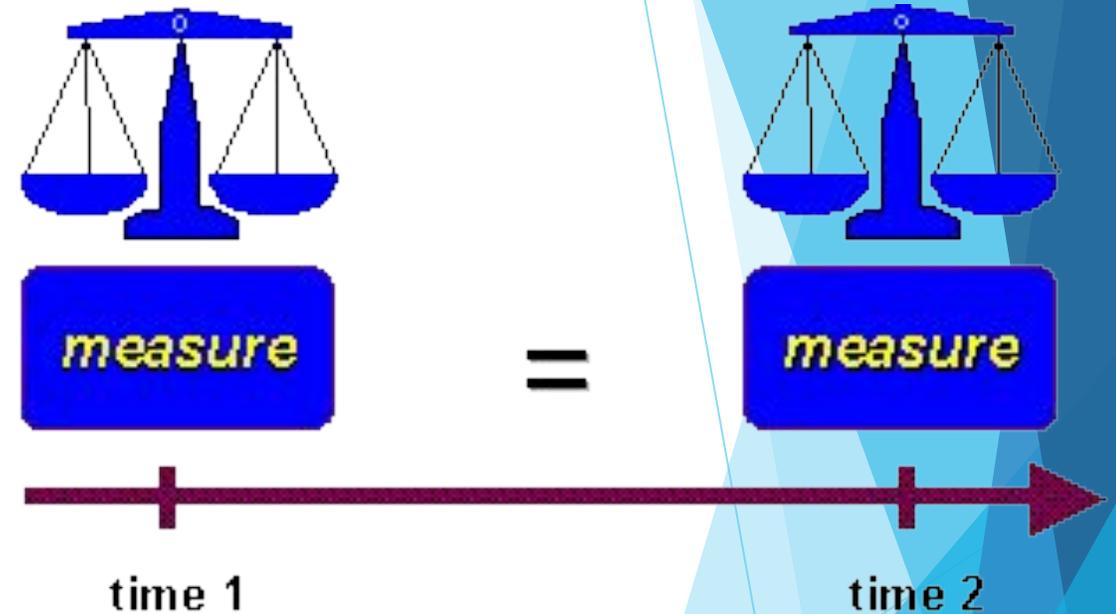
Will the test yield the same results if the test papers are marked:

- ▶ by two or more different examiners
- ▶ the same examiner on different occasions?



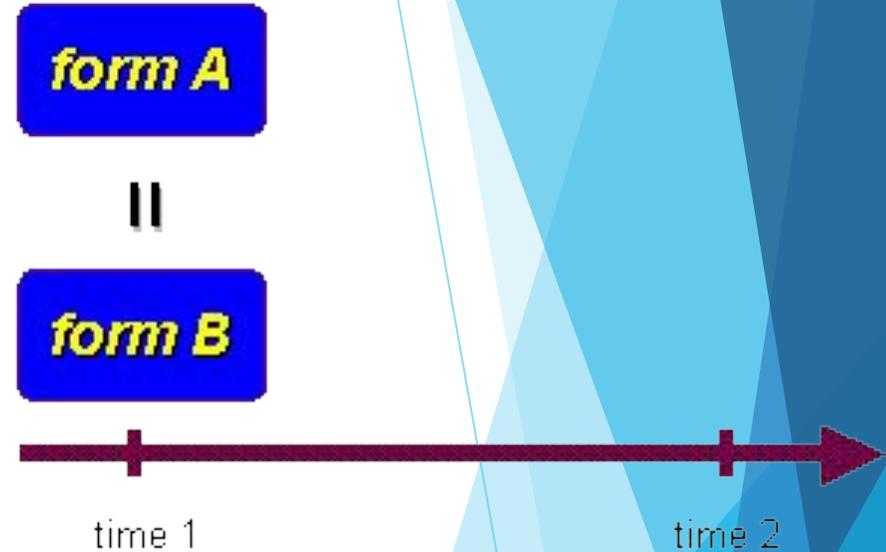
# Test - Retest reliability

- ▶ Repeatability of test scores with the passage of time
- ▶ RR reliability is assessed when same test is given to the same sample of learners on different occasions with no or little instruction in between
- ▶ Based on the assumption that constructs are more or less stable



# Parallel-Form Reliability

- ▶ Parallel form reliability indicates how consistent test scores are likely to be if a person takes two or more forms of a test
- ▶ Two parallel forms of test should measure the construct equally well
- ▶ For a reliable test, there is no difference which form of the test (A or B) the person takes



# Fluctuations in the learner

Factors beyond the control of the test designer:

- ▶ Sickness
- ▶ Fatigue
- ▶ No sleep on the night before the test or just a “bad day”
- ▶ Emotional problems



# How to balance between validity and reliability?

- ▶ It is possible to design a very valid communicative test which is not reliable (scorer reliability).
- ▶ Multiple-choice questions are one way to ensure that a test is more reliable, but is it valid to test speaking or writing?
- ▶ The key principles of validity and reliability need to be weighed up against each other when we design a test.

# 3. Practicality

Tests need to be TEACHER-FRIENDLY,  
i.e. they need to be:

- ✓ ...within the means of financial limitations;
- ✓ ...within time constraints;
- ✓ ...easy to administer, score and interpret

Thus...

# IMPRACTICAL!!!

- ▶ ... a test which is prohibitively **expensive**
- ▶ ...a test of language proficiency that would take students **10 hours to complete**
- ▶ ...a speaking test that requires individual **10 minutes one-to-one talk** for a group of 50 test-takers and only one scorer;
- ▶ .....a test that takes students **a few minutes to complete** and **several hours for the examiner to prepare and/or correct**
- ▶ ...a test which **can be scored only by computer** in a location without easy access to computers and internet connection

## 4. Washback

- ▶ Effect and consequences of a test on S, S's parents, Ts, schools, administrations, employers etc.
- ▶ Can have a positive or negative impact on the teaching and learning process

# Examples of positive washback

<b>On learners</b>	<i>• Provide a qualification</i>
	<i>• Provide motivation</i>
	<i>• Serve as a revision tool</i>
	<i>• Provide feedback</i>
<b>On teachers</b>	<i>• Identify struggling learners in a class</i>
	<i>• Diagnose common learner errors to modify instruction</i>
<b>On teaching institutions and schools</b>	<i>• Increase accountability of school</i>
	<i>• Identify weaknesses of a syllabus</i>
	<i>• Encourage a balanced curriculum</i>

# Possible negative washback

- ▶ Preparation for a test may take up teaching time.
- ▶ A test can be used as a way for teachers to exert their authority.
- ▶ Learners only practice the things that they know will be in the test, and ignore everything else.
- ▶ Learners feel stressed or nervous about the test conditions, the results and their image.
- ▶ Learners feel demotivated either by the prospect of revising for the test or at the thought of getting low marks.
- ▶ The way the test is marked may penalize errors rather than give credit for what the learner has done correctly.
- ▶ Test results may cause a feeling of divisions within the class.
- ▶ Improving test results can seem more important than learning - this often means that the range of skills taught becomes narrower.

# 5. Fairness

- ▶ For a test to be **fair** it should not discriminate against any subgroups of test takers or give advantage to other groups.
- ▶ It should also be fair to those who rely on the results.

## 6. Authenticity

- ▶ Our aim is to prepare students to function in the real world.
- ▶ Assessment should mirror real world situations and contexts
  - ▶ formats and tasks
  - ▶ authentic use of target language
- ▶ Authenticity is motivating!

## 7. Transparency

- ▶ Availability of information about assessment
- ▶ Information should include:
  - ▶ what they have to do to succeed, outcomes
  - ▶ expected content and format
  - ▶ time allocated for task, deadlines
  - ▶ Weighing of items or sections
  - ▶ grading criteria
  - ▶ useful feedback for improvement

## 8. Security

- ▶ **Students:**
  - ▶ Cheating, “collaborative” test-taking, plagiarism or any other kind of intellectual dishonesty is forbidden
- ▶ **Staff:**
  - ▶ There are clear security guidelines for all stages of assessment that must be followed
- ▶ **There are severe consequences for breaches of security.**