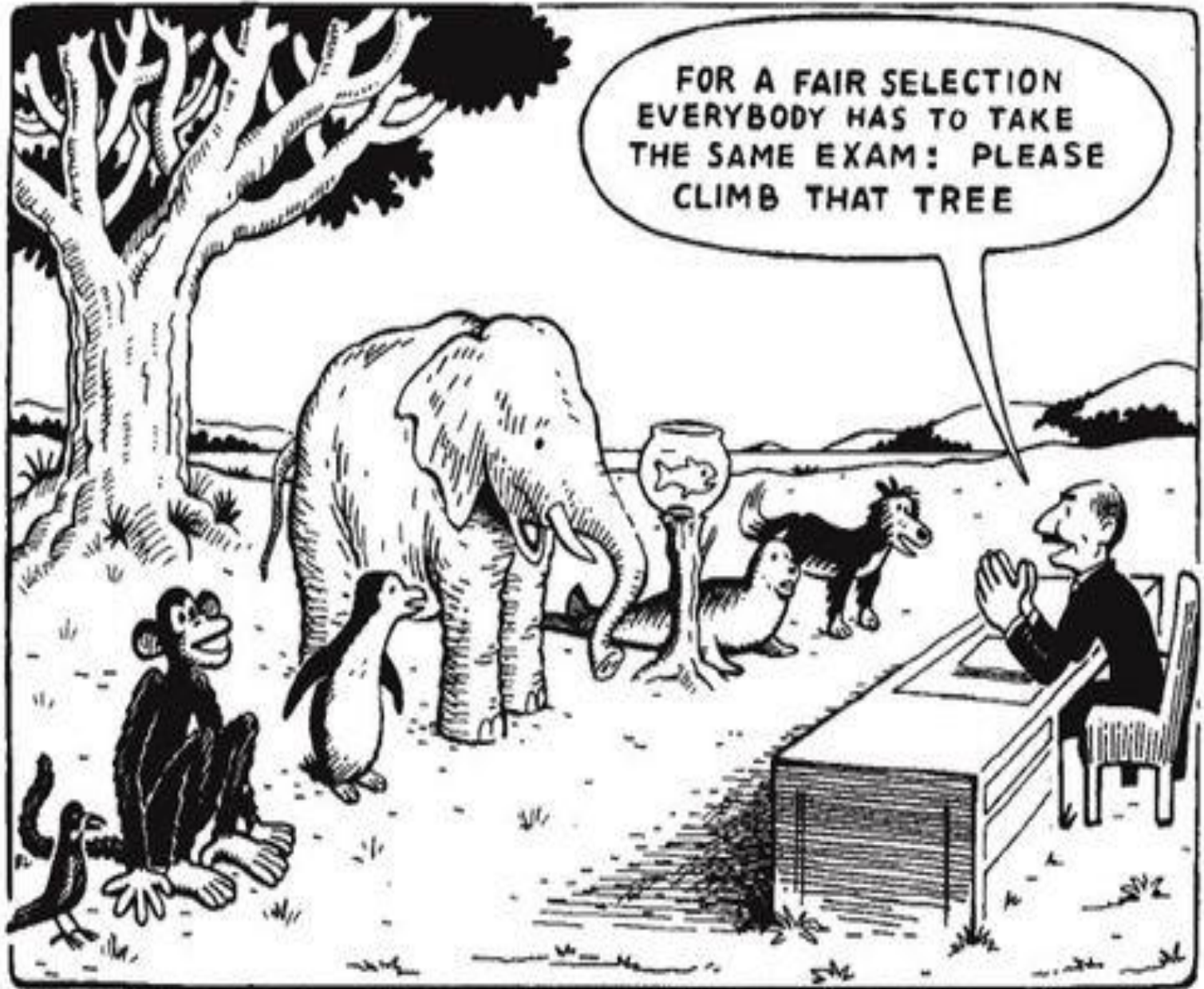


TEA: **Testing, Evaluation, Assessment**

Key concepts

Elena Nadtocheva,
Yekaterinburg

FOR A FAIR SELECTION
EVERYBODY HAS TO TAKE
THE SAME EXAM: PLEASE
CLIMB THAT TREE



Testing, Evaluation, Assessment

- Testing

A. the systematic collection, analysis, and interpretation of data for the purpose of determining the value of and decision-making about a program or policy; it looks at what we have set out to do, what we have accomplished, and how we accomplished it.

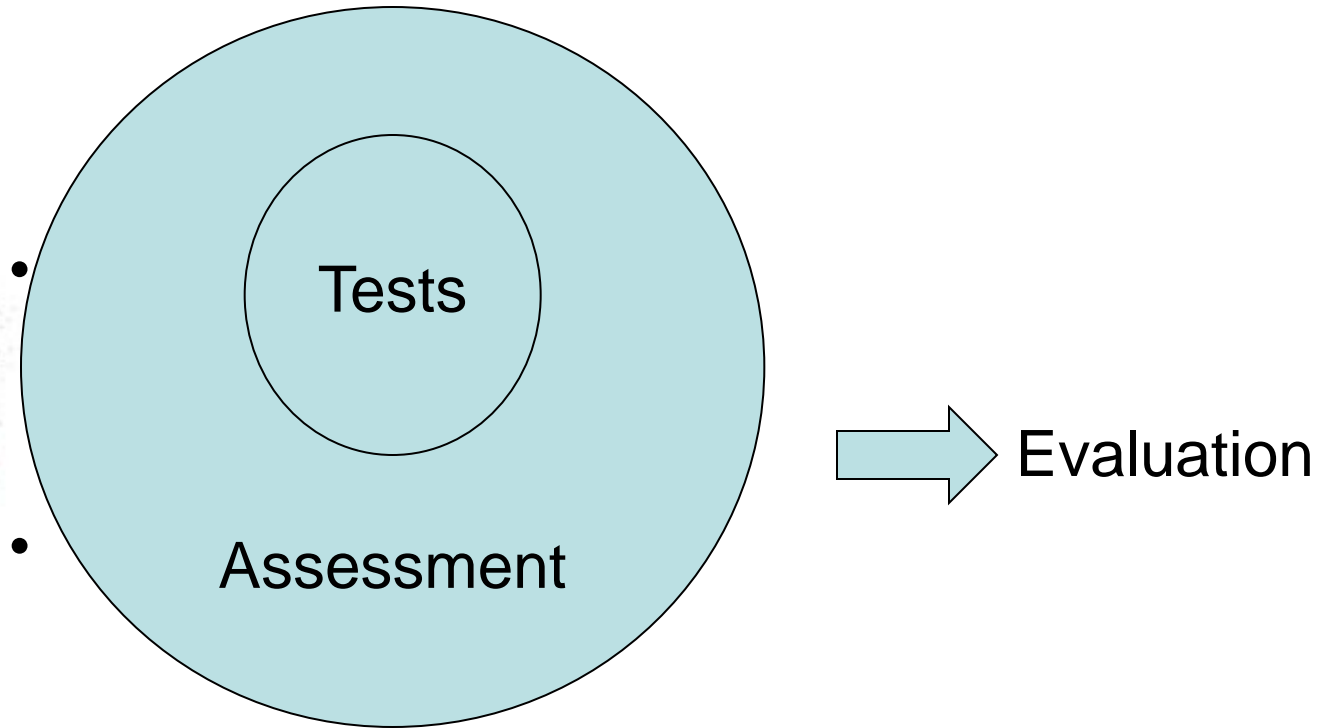
- Evaluation

B. A systematic approach to collecting information and making inferences about the ability of a student or the quality or success of a teaching course on the basis of various sources of evidence. May be done by test, interview, questionnaire, observation, etc.

- Assessment

C. Any procedure for measuring ability, knowledge, or performance.

TEA



- H.Douglas Brown. Language Assessment. Pearson, 2010

Testing

In each of the items below, select the definition that correctly defines the word. You have 2 minutes to complete the test.

1. Onager

- a. a phrase or word that quantifies a noun
- b. an adolescent approaching puberty
- c. an Asian wild ass with a broad dorsal stripe

2. Shroff

- a. a fragment of an ancient manuscript
- b. a banker or money changer who evaluates coins
- c. (archaic) past tense form of the verb *to shrive*

3. Chary

- a. overly talkative
- b. discreetly cautious about dangers and risks
- c. optimistic, positive, looking on bright side

Testing

In each of the items below, select the definition that correctly defines the word. You have 1 minute to complete the test.

1. *Onager*

- a. a phrase or word that quantifies a noun
- b. an adolescent approaching puberty
- c. **an Asian wild ass with a broad dorsal stripe**

2. *Shroff*

- a. a fragment of an ancient manuscript
- b. **a banker or money changer who evaluates coins**
- c. (archaic) past tense form of the verb *to shrive*

3. *Chary*

- a. overly talkative
- b. **discreetly cautious about dangers and risks**
- c. optimistic, positive, looking on bright side

Test Design



Test Design

- What?
- Why?



TEST TYPES

- ✓ Achievement tests
- ✓ Diagnostic tests
- ✓ Placement tests
- ✓ Proficiency tests
- ✓ Aptitude tests



TEST TYPES

Which tests are designed/used to identify learners' strengths and weaknesses in the target language?

- ✓ Achievement tests
- ✓ Diagnostic tests
- ✓ Placement tests
- ✓ Proficiency tests
- ✓ Aptitude tests

TEST TYPES

Which tests are designed/used to identify learners' strengths and weaknesses in the target language?

- ✓ Diagnostic tests



TEST TYPES

Which tests are administered during and at the end of a course?

- ✓ Achievement tests
- ✓ Diagnostic tests
- ✓ Placement tests
- ✓ Proficiency tests
- ✓ Aptitude tests

TEST TYPES

Which tests are administered during and at the end of a course?

- ✓ Achievement tests



TEST TYPES

- Which tests are designed to measure a person's capacity for language learning?
 - ✓ Achievement tests
 - ✓ Diagnostic tests
 - ✓ Placement tests
 - ✓ Proficiency tests
 - ✓ Aptitude tests



TEST TYPES

- Which tests are designed to measure a person's capacity for language learning?
- ✓ Aptitude tests



TEST TYPES

Content of these tests is based on specifications of what test taker is able to do to be considered 'proficient' for a particular purpose.

- ✓ Achievement tests
- ✓ Diagnostic tests
- ✓ Placement tests
- ✓ Proficiency tests
- ✓ Aptitude tests

TEST TYPES

Content of these tests is based on specifications of what test taker is able to do to be considered 'proficient' for a particular purpose.

- ✓ Proficiency tests



TEST TYPES

- Which tests are used to determine where a learner should begin within a level system in an institution?
- ✓ Achievement tests
- ✓ Diagnostic tests
- ✓ Placement tests
- ✓ Proficiency tests
- ✓ Aptitude tests

TEST TYPES

- Which tests are used to determine where a learner should begin within a level system in an institution?
- ✓ Placement tests




Quiz on Test Purposes

Which type of a test would be most appropriate in the following situations?



Quiz: Which type of a test would be most appropriate in the following situations?

- An army unit which needs to choose five people for intensive language training courses.
 - – Aptitude test
- A group of doctors who are going to work in the USA for one year and are going to study together before they go.
 - – Diagnostic test
- Multilingual post-graduates who need to know their level of English before applying to Australian universities.



Quiz: Which type of a test would be most appropriate in the following situations?

- An army unit which needs to choose five people for intensive language training courses.
- A group of doctors who are going to work in the USA for one year and are going to study together before they go.
- Multilingual post-graduates who need to know their level of English before applying to Australian universities.
- A class of school children who need to be tested at the end of the school semester.
- A group of engineers from different companies who are going to study an ESP course available at 3 different levels.

Quiz: Which type of a test would be most appropriate in the following situations?

- An army unit which needs to choose five people for intensive language training courses.
 - – Aptitude test
- A group of doctors who are going to work in the USA for one year and are going to study together before they go.
 - – Diagnostic test
- Multilingual post-graduates who need to know their level of English before applying to Australian universities.
 - – Proficiency test
- A class of school children who need to be tested at the end of the school semester.
 - – Achievement test
- A group of engineers from different companies who are going to study an ESP course available at 3 different levels.
 - – Placement test

TESTS

✓ What?



What Is Tested?

- You will hear a conversation between a senior manager, called Sue, and her assistant, called David.

For each question 23 – 30, mark one letter (A, B or C) for the correct answer.

You will hear the conversation twice.

What Is Tested?

You are going to attend an engineering exhibition in Frankfurt soon.

Write a memo to your assistant:

- explaining why you will be away
- letting her know the dates you will be away
- saying what work she should do while you are away.

Write 30 – 40 words.

What Is Tested?

- Look at questions 1 – 5.
- In each question, which sentence is correct?
- For each question, mark one letter (A, B or C) on your Answer Sheet.

Office Staff Required

Experience essential

Full training given (leading to recognised qualifications)

1. Applicants must have
 - A relevant qualifications.
 - B previous experience.
 - C recognised training

Major Principles of Language Assessment

- ✓ Validity
- ✓ Reliability
- ✓ Practicality
- ✓ Impact



Validity

- Construct validity
- Content validity
- Face validity
- Concurrent Validity
- Predictive Validity



Construct validity

- How well test measures what it is supposed to measure?

E.g. Student's oral fluency:

- speed
- rhythm
- juncture
- (lack of) hesitations
- etc.

Content validity

Test what you have taught!

- Content relevance
- Content coverage

Bachman 1990: 244.



Face validity

- Test is testing what it is supposed to test.
- Looks relevant and worthwhile investing time in.



Concurrent validity

- Statistically establishes closeness of a new test to its criterion.
- A test has it if its results are supported by other concurrent performance beyond the assessment itself.



Predictive Validity

- Does the test assess a test-taker's likelihood of future success?

Ready

- to 'move on'?
- to cope with linguistic demands in real-life situations?



Reliability

- Inter-rater reliability
- Intra-rater reliability
- Test-retest reliability
- Sampling



Reliability

- Inter-rater reliability:

Several raters have stable scores of the same test.

- Intra-rater reliability

An examiner is consistent, his judgement is objective at all times.

Reliability

- Test-retest reliability

A test should produce consistent results when it's done several times under the same conditions.

- Sampling

Choosing a selection of tasks from the domain of language under test. The test is then said to contain a sample of all possible items and tasks that could be included.

Practicality

- ✓ Test security
- ✓ Test delivery
- ✓ Test administration
- ✓ Standardisation and marking



Practicality

✓ Test security

Protection against theft, illegal distribution, etc.

✓ Test delivery

The way it is delivered (paper, online, etc.)



Practicality

✓ Test administration

Where? How Many? How to deliver?

✓ Standardisation and marking

A universally accepted system of marking



Impact

- Washback
- Stakeholders
 - Context
- Test preparation
- Tests as political instruments



Washback

- How tests influence both teaching and learning.
- A test should provide beneficial feedback.



Question Time



Thank you!

elenanadtocheva@yandex.ru



elta-urals