

LANGUAGE TESTING: KNOWN AND UNKNOWN (notes on the workshop)

ACKNOWLEDGEMENTS

The idea of the workshop was prompted by a series of training sessions held within the project "Assessment in ELT" sponsored by the British Council Ukraine. In this respect we would like to thank our trainers – **Dr Charles Boyle**, Delaso UK Ltd Director, **Zoltán Tószegi**, Examinations Manager, and **Patricia Aresvik**, Senior Training Consultant, who provided our Project team both inspiration and criticism as we worked through numerous assessment issues.

INTRODUCTION

Testing at large has recently become one of the most burning issues that Ukrainian educational community keeps discussing. With introduction of credit modular transfer system in universities and the perspective of introducing independent testing of secondary school leavers' foreign language skills, the issues of *language testing* have become the area of special interest and concern. So, the overall **OBJECTIVE** of the workshop is to make the readers reflect on their knowledge and experience in language testing as well as brief them on the following testing issues:

- concepts underpinning language testing;
- qualities that make the language test a reliable measurement instrument;
- task types which are most commonly used in language testing.

Another point of concern for us is to introduce the **key terminology** used in language testing. We expect this aspect to be of relevant importance for our readers for some reasons. First of all, *concepts* encoded in terms speak volumes. Another point is the ability to *name* appropriately what we come across and use (quite often intuitively rather than consciously) on a daily basis. The expected outcome of our terminology awareness is our ability to speak common testing language with our colleagues both domestically and abroad. This is becoming tremendously significant with our ambitions to join the European language testing community and adopt European language testing practices.

Finally, we plan to lead the workshop participants through some **practical issues** of test (test task) design. The main focus in this respect is to demonstrate that there are no "trifles" in language testing – every single detail of the test/task **DOES** make real difference! Thus, we designed every task in the way that can, on the one hand, give the workshop participants food for thought on both – the layout and the content of the test task, and, on the other hand, enable us to make adequate comments on all that and thus update our trainees on key testing issues.

FOOD FOR THOUGHT: Before you start doing any of the tasks in each section (STEP) analyse the task itself. Please consider the points that follow.

1. Are you clear about:
 - *what* you are supposed to do;
 - *how* to perform the task (tick, circle, cross out – whatever?);
 - *where* to record your responses (is this mentioned in the rubric and is special space provided?);
 - *how many* points you can get for each item as well as for the task as a whole (is this mentioned in the rubric and is the total score provided?);
 - *what* can help you to do the task (is the example provided? is it italicized?);
 - *which* task type this task falls into.
2. What do you think of the *language* of the rubric and the items (is it appropriate in terms of its accuracy, complexity, style)?
3. Is the *layout* of the task user friendly (consider its type, size, intervals – whatever)?

STEP 1: WHY DO WE TEST?

Read these comments from teachers. Decide why they are giving a test in each case. For each question choose the correct answer (A, B, or C). You can also provide your own answer (D).

- "Right! As you can't be quiet and won't do what I tell you, we'll have a test!"
 - A. to place students in a group according to their knowledge
 - B. to measure students' potential for learning the language
 - C. to calm students down and kill time during the lesson
 - D. ???
- "They never do any homework, so I give them lots of short tests instead."
 - A. to discover students' specific strengths and weaknesses
 - B. to find out what motivates students to learn foreign language
 - C. to revenge on students and stimulate them to do homework
 - D. ???
- "I haven't got enough grades in the register for class 10B, so I'll give them a test next lesson."
 - A. to measure students' ability to operate in an academic context
 - B. to assess students' progress up to a certain point of the course
 - C. to make sure you look a strict and very demanding teacher
 - D. ???
- "What do you mean I didn't tell you we were going to have a test? It shouldn't matter. You should be prepared anyway – it's on material we've just done!"
 - A. to assess students' progress in relation to a particular textbook
 - B. to discover strengths and weaknesses of your teaching style
 - C. to pretend you're concerned with your students' progress
 - D. ???
- "Now we'll find out what you don't know. I've got a nice, tricky, little test for you!"
 - A. to make a decision on students' admission to school
 - B. to assess students' general language abilities and skills
 - C. to make students believe in your abilities as a test designer
 - D. ???

FOOD FOR THOUGHT: Analyse the testing situations described above. Consider the following points:

- what types of tests we as teachers use for classroom activities or while assessing students' knowledge and skills?
- do the tests we use pursue appropriate aims?
- does any of the tests teachers mentioned above correspond to the types of tests described below? Do the tests *you* use fall into these types of tests?

achievement test A test designed to measure the extent of learning achieved by a candidate in relation to a particular course, textbook, etc., i.e. a curriculum-dependent test. (Also referred to as *attainment test*.)

aptitude test A test designed to predict or measure a candidate's potential for success within a particular area of learning, i.e. in learning a foreign language, or on a specific course of study.

diagnostic test A test which is used for the purpose of discovering a learner's specific strengths and weaknesses. The results may be used in making decisions on future training, learning or teaching.

entrance test A test used for determining whether or not a candidate gains admission to a particular institution or course of study.

placement test A test administered in order to place students in a group or class at a level appropriate to their degree of knowledge and ability.

proficiency test A test which measures general ability or skill, without reference to any specific course of study or set of materials.

progress test A test administered part of the way through a course of instruction, in order to assess learning up to that point.

specific purpose test A test designed to measure a candidate's ability to operate in a specific professional or academic context. The content of the test is designed accordingly, on the basis of an analysis of the language tasks candidates will need to deal with in the target language use situation.

STEP 2: WHAT DO WE MOST COMMONLY MISUNDERSTAND ABOUT LANGUAGE TESTING

FOOD FOR THOUGHT:

- Consider the following ideas that L.F Bachman and A. Palmer identified as "**the most common misconceptions of language testing**" (Bachman & Palmer, 2000).
- Think over the problems that might arise as a result.
- Have *you* ever come up with the same ideas about testing?

MISCONCEPTION 1. Believing that there is one "best" test for any given situation. As language teachers (since most commonly we ARE language teachers, not professional testers) we DO believe that if we follow the model of a test that has been designed and developed by the "expert" in the field and the test was widely recognized and used, it would automatically be useful for our particular needs.

- **Resulting problems:** tests which are inappropriate for the test takers.

MISCONCEPTION 2. Misunderstanding the nature of language testing and language test development. We rarely consider correlation between language testing, language teaching and language use, as well the dichotomy *language testing :: language learning* with respect to the former having laws and regulations which might differ from those of the latter. We do realise that cognitive processes involved in language use and, correspondingly, language learning are different for different learners and vary with different language activities. But we rarely (if ever) doubt whether similar testing procedures require similar cognitive efforts and strategies from different individuals whom we have for our target population. So we practically never analyse if a single model would provide the most suitable test for the variety of our particular test-takers, particular uses and areas of language ability that are to be measured in our particular situation.

- **Resulting problems:** tests which do not meet the specific needs of the test users.

MISCONCEPTIONS 3-4. Placing blind faith in the technology of measurement as well as having unreasonable expectations about what language tests can do and what they should be. We often expect some "expert" to offer us some ready-made recipe of an "ideal" language test and/or teach us how to create a "good" test. But the very concept of a "good" language test is still vague and abstract enough for many of us, since we have no idea of what makes a really "good" test – its layout? the text selected? the task format chosen? the correlation between the task format and the skill we intend to measure? the language of the rubric – whatever?!

- **Resulting problems:**
 - uninformed use of tests or testing techniques simply because they have become popular;
 - becoming frustrated when one is unable to find or develop the "perfect" test;
 - loss of faith in one's own capacity for developing and using tests appropriately, as well as a feeling that language testing is something that only "experts" can understand and do;
 - being placed in a situation of trying to defend the indefensible, since many students, as well as administrators, have unreasonable expectations.

STEP 3: TERMINOLOGY USED IN LANGUAGE TESTING

TESTING TERMINOLOGY QUIZ (1)

You are going to consider concepts related to testing. In each group of testing terms one word is odd. Write the odd words in the boxes below. An example (0) has been done for you.

- 0 *evaluation, assessment, testing, scoring*
- 1 trialling, administration, marking, revision
- 2 test taker, test user, testee, candidate
- 3 interlocutor, assessor, marker, examiner
- 4 paper, task, item, band
- 5 item type, task type, test method, test construct
- 6 matching, stem, option, distractor
- 7 rubric, specification, instruction, question
- 8 feedback, washback, impact, backwash
- 9 score, point, scale, grade
- 10 cut-off score, bias, halo effect, boundary effect

Write your answers here:

0	<i>Scoring</i>
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

Total Score – 10 points

TESTING TERMINOLOGY QUIZ (1): COMMENTS

0 *evaluation, assessment, testing* – all the terms stand for the concepts related to measuring language skills and knowledge, cf.:

assessment In language testing, the measurement of one or more aspects of language proficiency, by means of some form of test or procedure.

evaluation Gathering information with the intention of using it as a basis for decision-making. In language testing, evaluation may focus on the effectiveness or impact of a programme of instruction, examination, or project.

testing a method of measuring a person's ability or knowledge in a given domain

1 *trialling, administration, revision* – stages of a testing cycle, cf.:

trialling A stage in the development of test tasks aimed at ascertaining whether the test functions as expected.

administration The date or period during which a test takes place. Many tests have a fixed date of administration, while others may be administered on demand.

revision A stage in the development of the test that is aimed at modifying test tasks.

2 *test taker, testee, candidate* – synonymous nominata that identify those who take a test.

3 *interlocutor, assessor, examiner* – people involved in testing speaking, cf.:

assessor / examiner Someone who assigns a score to a candidate's performance in a test, using subjective judgment to do so. In oral testing the roles of assessor and interlocutor are distinguished. Also referred to as examiner or rater.

interlocutor In a test of speaking, the examiner who explains the tasks, asks questions and generally interacts orally with the candidate(s). The interlocutor may also make the assessment of candidate(s) and assign scores, or this may be done by a second examiner, who observes but does not interact with the candidate(s).

4 *paper, task, item* – structural components of a test, cf.:

paper A way of referring to a test component, e.g. Reading Paper, Listening Paper.

task A combination of rubric, input and response. For example, a reading text with several multiple choice items, all of which can be responded to by referring to a single rubric.

item Each testing point in a test which is given a separate mark or marks. Examples are: one gap in a gap-filling test, one multiple choice question with three or four options; one sentence for grammatical transformation; one question to which a three-word-length response is expected.

5 *item type, task type, test method* – synonymic descriptive terms pertaining to a variety of techniques used in testing,

item type Test items are referred to by names which tend to be descriptive of the form they take. Some examples are: multiple choice, sentence transformation, short answer, etc.

task type Test items are referred to by names which tend to be descriptive of the form they take. Some examples are: multiple choice, sentence transformation, short answer, etc.

test method Language ability can be tested using a variety of methods such as multiple choice, composition, oral interview, etc. Test method has been observed to interact with ability in the measurement of performance.

6 *stem, option, distractor* – terms most commonly related to describing components of a multiple choice task, cf.:

stem Part of a written prompt, usually an incomplete sentence for which the completion must be supplied from the selected options.

option(s) The range of possibilities in a multiple choice item or matching task from which the correct one must be selected.

distractor Each incorrect option in a multiple choice item.
Compare: key response.

7 rubric, instruction, question – guidelines for testees on what to do while performing a test, cf.:

rubric The instruction given to candidates to guide their responses to a particular test task.

instruction(s) General directions given to test candidates, for example on the front page of the answer paper of booklet, giving information about such things as how long the test lasts, how many tasks to attempt and where to record their responses.

question Sometimes used to refer to a test task or item.
See: item, task.

8 washback, impact, backwash – cf.:

impact The effect created by a test, both in terms of influence on general social and educational processes (*Macro-impact*), and in terms of the individuals who are effected by test results (*Micro-impact*).

backwash The impact of a test on classroom teaching. Teachers may be influenced by the knowledge that their students are planning to take a certain test, and adapt their methodology and the content of lessons to reflect the demands of the test. The result may be positive or negative. Also referred to as **washback**.

9 score, point, grade – expressions of test results, cf.:

score The total number of **points** someone achieves in a test. **See also:** **scoring** – calculating and assigning the total number of points someone achieves in a test.

point Numerical expression of candidate's performance on each individual item.

grade A test score may be reported to the candidate as a grade, for example on a scale of A to E, where A is the highest grade available, B is good pass, C a pass, and D and E are failing grades.

10 bias, halo effect, boundary effect – constraints that the test can come across on its way to be a reliable measurement instrument, cf.:

bias A test or item can be considered to be **biased** if one particular section of the candidate population is advantaged by some feature of the test or item which is not relevant to what is being measured. Sources of bias may be connected with gender, age, culture, etc.

halo effect A tendency of assessors involved in subjective examining to be influenced by a candidate's performance on certain test tasks and consequently to give too high or low score for another task.

boundary effect(s) The effect(s) of a test's being too easy or too difficult for a particular group of candidates. Scores tend to accumulate at the top of the distribution (*ceiling effect*) or at the bottom (*floor effect*).

TESTING TERMINOLOGY QUIZ (2)

Match the testing terms in the box to each of the definitions below. Use each **ONCE** only. There are three additional words/phrases you do not need. An example (0) has been done for you.

marking scoring descriptor marker test construct
double marking stakeholder feedback scale band
cut-off score test user competence specification

0	Comments of people involved in the testing process (examinees, administration, etc.) which provide a basis for evaluating that process. This may be gathered informally, or using specially-designed questionnaires.	
1	Assigning a mark to a candidate's responses to a test. This may involve professional judgement, or the application of a mark scheme which lists all acceptable responses.	feedback
2	Someone who assigns a score to a candidate's responses to a written test. This may involve the use of expert judgement, or, in the case of a clerical marker, the relatively unskilled application of a mark scheme.	
3	A set of numbers or categories for measuring something.	
4	The minimum score a candidate has to achieve in order to get a given grade in a test or an examination.	
5	In its broadest sense, part of a scale. In an item-based test this covers a range of scores which may be reported as a grade. In a rating scale designed to assess a specific ability such as speaking or writing, this normally represents a particular level.	
6	Calculating the total number of points someone achieves in a test.	
7	Procedure during which a candidate's performance is rated by two examiners separately and the marks are later compared.	
8	The person or institution (e.g. a teacher or employer) using the results of a test in order to make a decision concerning the test taker.	
9	A detailed set of documentation normally drawn up during the process of designing a new test or revising an existing one. This gives details of what the test is designed to measure and how it will be tested, i.e. the design, content, level, task and item types used, target population, use of the test, etc. This also often includes specimen materials.	
10	Those aspects of the candidate's underlying knowledge or skill which are the target of measurement in a test.	

Total Score – 10 points

STEP 4: TASK / ITEM TYPES

FOOD FOR THOUGHT:

- Every time you set yourself on designing a test (test task), think of what language skills/subskills/operations you are going to assess and what task types are better suited for the purpose.
- To make sure you can identify different task/item types do the **MATCHING** tasks below.

TASK / ITEM TYPES QUIZ (1)

Match terms (1-14) with their definitions (a-m). There are two terms that you do not need. Write your answers in the boxes below. An example (0) has been done for you.

0 summary

- | | |
|--------------------------|-----------------------|
| 1 information transfer | 8 sequencing |
| 2 problem solving | 9 information gap |
| 3 dictation | 10 word formation |
| 4 multiple matching | 11 sentence expansion |
| 5 True/False | 12 multiple choice |
| 6 transformation | 13 proof reading |
| 7 short answer questions | 14 c-test |

- a This task type involves taking information given in a certain form and presenting it in a different form (e.g. labelling a diagram).
- b An item type where the candidate has to produce a form of a word based on another form of the same word which is given as input e.g. *This kind of work requires an of technical vocabulary.* (UNDERSTAND)
- c A type of selected response item where the candidate has to indicate whether a series of statements are correct or incorrect in relation to a text.
- d An open-ended item for which the candidate is required to formulate a written answer using a word or a phrase.
- e An item type in which a prompt in the form of a string of content words is given, but prepositions, auxiliary verbs, articles, etc. are absent. The candidate's task is to add the missing words, thus turning the prompt into a complete grammatical sentence.
- f In this task the candidate is required to listen to a text and write down the words which are heard.
- g A test task which involves checking a text for errors of a specified type, e.g. spelling or structure.
- h A type of test item which consists of a question or incomplete sentence, with a choice of answers or ways of completing the sentence. The candidate's task is to choose the correct option from a set of three, four or five possibilities.
- i A technique of language testing which simulates real communication by setting up situations in which students do not share the same information, and thus need to communicate with each other in order to complete a task.
- j A technique in which two sets of stimuli have to be brought together e.g. paragraph headings and their corresponding paragraphs.
- k An item type in which candidates are asked to rewrite or complete sentences so that the second sentence remains faithful in meaning to the original, e.g. the first sentence may be active, and the candidate's task is to present the identical content in passive form.

- l A type of gap-filling task in which the second half of certain words is deleted. The frequency of deletion can be as high as every second word. The candidate's task is to complete the partially deleted words.
- m A task in which candidates are asked to produce a precis or shortened version of a text or part of a written or spoken text.

Write your answers here:

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
m														

Total Score – 12 points

TASK / ITEM TYPES QUIZ (2)

Match the item types listed below to the examples (1-30). Write your answers in the boxes. An example (0) has been done for you.

NB! Some examples are represented by both the rubric and a fragment of the input. In case the rubric is too explicit, or the input is too extensive to be printed here, you will have to judge by one of the components only.

- Matching headings to paragraphs
- Error detection
- Dictation
- Gapped text (clauses)
- Word-ordering/sequencing/unscrambling
- Essay
- Matching banked words to definitions
- Gap-filling (paragraphs)
- Multiple choice
- Banked gap-filling (words)
- Proofreading
- Note taking
- Translation
- Open gap filling
- Role play
- True/False
- Multiple Matching (questions to texts)
- Short Answer Questions
- Register transfer
- Summary
- Dialogue/ discourse completion
- Note expansion
- Word associates
- Punctuation
- Text Ordering/Sequencing
- Word formation
- Phonemic discrimination
- Sentence transformation/Paraphrasing
- Listing
- Table completion
- Multiple matching (words to pictures)

(0)

Cross out the unnecessary words.

- 1 I need the book which I lent it to you.
- 2 She was too much tired to go out with us.

0	Error detection
---	-----------------

(1)

- 1 British Government – recently started – will participate in project – in co-operation with Scandinavia – look into problem of acid rain.

2 Experts think – unnecessary – time on project – expensive – already have solution.

1	
---	--

(2)

Вони часто гуляли містом і розмовляли про все на світі. Ось звідки появилось те дивне відчуття. Відчуття, що ти будеш оберігати цю людину, ніколи не будеш сперечатися з нею через дрібниці. Невже можна з самого початку все зіпсувати? Невже таке почуття може зійти нанівець? Вона згадувала, і десь всередині приємно розливалось тепло, від якого розцвітав бузок і яким горіли його ніжні, напрочуд ясні очі...

2	
---	--

(3)

1. And /his /the /friends /Tom /go /the /end /will /to /week /country /for.
2. Heard /suddenly /some /to /from /I /the /noise /hurried /door /outside /and.
3. Crowds /everywhere /people /were /there /of.

3	
---	--

(4)

Match the crimes in the picture with words from the list below.

arson assault blackmail burglary fraud hijacking kidnapping mugging murder robbery shoplifting theft

4	
---	--

(5)

Complete gaps 1-15 in the text below.

You'll lose a year of your life looking for missing possessions

The average person in Britain spends a year of their life looking for lost items, with Mondays being the most likely day (0) to misplace vital possessions.

Research published today shows that (1).....to popular belief, technological advances have (2).....helped people find, file and organize things (3) easily but (4)increasing the likelihood (5).....people losing money, handbags, wedding rings and (6).....important possessions.

5	
---	--

(6)

You are going to read a magazine article about house prices and transport connections. Eight paragraphs have been removed from the article. Choose from the paragraphs (A-I) the one which fits each gap (14-20). There is one extra paragraph which you do not need to use. There is an example at the beginning (0).

6	
---	--

(7)

- A Carl talked to Ray in a pub about money problems.
- B Ray suggested robbing a bank.
- C Carl and Ray spent months planning to rob a bank.
- D Carl drove to a bank in a small town in Scotland.
- E Ray came out of the bank with his hands in the air.
- F Carl was about to drive off when the police arrested him.
- G Carl was sentenced to twelve years in prison.
- H The heavy prison door shut behind Carl.

7	
---	--

(8)

You are going to hear a radio interview with a woman who is talking about her life. First you will have 30 seconds to study the questions below. Then you will hear the text twice. While listening, answer the questions (1-9) in a maximum of FOUR words. There is one example (0) at the beginning. After the second listening, you will have 30 seconds to finalise your answers. Start studying the questions now.

Casey

0 How old is Casey's youngest child? 16 months.

8	
---	--

(9)

Reading Passage 3 has seven paragraphs A – G. Choose the most suitable headings for paragraphs A – G from the list of headings below. Write the appropriate number i-ix in boxes 28 – 33 on your answer sheet. The first one has been done for you as an example. There are more headings than paragraphs, so you will not use them all.

Paragraph Headings

- | | |
|-----------------------------|-----------------------------------|
| i) A Tragic Ending | vi) Reactions to Revolution |
| ii) A Life of Revolution | vii) A Life in Perspective |
| iii) Being Different | viii) The First Reaction to Burke |
| iv) Contradictory Behaviour | ix) Asserting the Rights of Women |
| v) The Work of Thomas Paine | |

9	
---	--

(10)

The marketing off the 250th anniversary of the Mozart's birth is well and truly underway. In the composers' home city of Salzburg, the elegant boutiques in the mane street is positively overflowing the Mozart souvenirs, from porcelain figurines to hand-made chocolates in the shaped of musical instruments.

10	
----	--

(11)

Read through an article on Leonardo da Vinci's Mona Lisa and write a 115-word summary of it.

11	
----	--

(12)

Complete each sentence exactly as it is said

At three o'clock, precisely, the guests were all assembled in the drawing room.

12	
----	--

(13)

New Neighbours

Card for candidate A	Card for Candidate B
<ul style="list-style-type: none"> ▪ Greet your new neighbour ▪ Introduce yourself ▪ Describe your old home ▪ Discuss why you like the new flat/neighbourhood ▪ Invite your neighbour to look at your room 	<ul style="list-style-type: none"> ▪ Greet your new neighbour ▪ Introduce yourself ▪ Describe your old home ▪ Discuss why you like the new flat/neighbourhood ▪ Invite your neighbour to look at your room

13	
----	--

(14)

Match the words (1—7) with the correct definitions (a-g).

- | | |
|-------------------------------|---|
| 1 to consult | a a doctor who is trained in general medicine |
| 2 a consultant | b medical treatment which involves cutting the body open |
| 3 a general practitioner (GP) | c the room in a hospital where operations are performed |
| 4 an operating theatre | d a senior hospital doctor who specialises in one area of medicine |
| 5 to prescribe | e to ask a doctor for information or advice |
| 6 surgery (uncountable) | f to tell a patient, usually in writing, what treatment or medicine to have |
| 7 a surgery (countable) | g a place where a doctor or dentist examines or treats patients |

14	
----	--

(15)

0 Sarita and Ranjin had known each other as children. T
1 Sarita's mother liked Ranjin a lot.

15	
----	--

(16)

Read the article in detail and list the arguments it contains for suggesting that 'male and female thought processes are different'.

16	
----	--

(17)

You are going to read a text about Claudia Schiffer. Some words are missing from the text. Choose the most appropriate word from the list (A-K) for each gap (1-10) in the text. There are two extra words that you don't need to use. Write your answers in the boxes after the text. There is an example at the beginning (0).

17	
----	--

(18)

The draft letter below is addressed to an unsuccessful candidate for the vacancy at *Slim Gyms*. Put a circle around any language that you think is too informal for the situation.

Dear Michael,

After our chat about the job, it's a pity but you didn't get it this time. On the day, the other people were better. The guys thought your resume and interview were fantastic, but you hadn't done enough for this job.

Anyway, now is a good time to say thanks for coming to *Slim Gyms* and we hope you find a job soon and have a good life.

Best wishes

18	
----	--

(19)

You will hear a recorded message on the employment hotline for Mail, UK. For questions 1-8, complete the notes about the type of job on offer.

Mail UK

Type of job: (1) drivers

We are looking for people who are:

good (2) - able to deal with all sort of people.

19	
----	--

(20)

The 1996 Paralympics, held in Atlanta, USA, were a 1) (demonstrate) of the best aspects of sporting 2)..... (compete).

20	
----	--

(21)

For questions 1-6, read the following text and then choose from the list A-J the best phrase given below it to fill each of the spaces. Each correct phrase may only be used once. Some of the suggested answers do not fit at all.

'Lump of rock' turns out to be world's first sculpture

New scientific data suggests (1) 220,000 years ago, 170,000 years earlier than previously thought. It is a discovery which could revolutionize our understanding of human development.

- A that the sculpture could stand upright
- B that probably took between 15 and 30 minutes to make
- C that the prehistoric object was intended to portray a human being

21	
----	--

(22)

- 1 notice, observe, watch, think
- 2 terrified, scared, thrilled, horrified
- 3 big, enormous, huge, tiny
- 4 effect, consequence, cause, result

22	
----	--

(23)

Complete the sentences using the words in bold.

- 1 It is too cloudy to go to the beach today
sunny Itto go to the beach today.

23	
----	--

(24)

- 1 According to the speaker, what do people like about teleworking?

- A You work fewer hours.
- B You don't have to commute.
- C You have more time for friends and family.

24	
----	--

(25)

Write a suitable phrase in each gap to complete this dialogue.

Mrs Taylor: Good afternoon, Mr Dee. I'm Mary Taylor. Pleased to meet you.

John Dee:²⁸, Mrs Taylor.

Mrs Taylor: Welcome to our company.²⁹
a good trip?

John Dee: Oh yes, it was fine thanks.

Mrs Taylor:³⁰ a cup of coffee before we start?

John Dee: Thanks very much, that would be nice.

Mrs Taylor:³¹ milk and sugar?

John Dee: Just milk please.

25	
----	--

(26)

Listen to the tape and circle the words you hear.

- batch – badge
- rich – ridge

26	
----	--

(27)

i was alone in the house reading a scary ghost story a s snow fell silently outside the only sound was the ticking of my old grandfather clock the dying fire cast an orange light onto the walls of my study

27	
----	--

(28)

You are going to read some information about restaurants in Edinburgh. For questions (21-35), choose from the restaurants (A-E). Some may be chosen more than once. When more than one answer is required, these may be given in any order. There is an example at the beginning (0).

28	
----	--

(29)

You are going to hear five people talking about these subjects. Complete this table with short answers to the questions.

	speaker	1	2	3	4	5
1	What have they had pierced, their ear(s) or their nose?					
2	When did they have it done?					
3					
		29				

(30)

You should spend about 40 minutes on this task. Present a written argument or case to an educated reader with no specialist knowledge of the following topic.

It has been claimed that workers over 50 are not responsive to rapidly changing ideas in the modern workplace and that for this reason younger workers are to be preferred. To what extent would you support or reject this idea?

You should write at least 250 words.

You should use your own ideas, knowledge and experience and support your argument with examples and relevant evidence.

30	
----	--

Total Score – 30 points

STEP 5. FOOD FOR THOUGHT (FINAL TASK)

Read the text about the nature of testing. On which points do you agree or disagree with its author? Give specific reasons to support your ideas.

A test should aim to fulfil three conditions.

It should test what the test writer wants it to test. If tests suffer from any of the faults you identified above it is

possible that this principle may be affected. Test validity presupposes the writer can be explicit about this and takes steps to ensure that the test reflects realistic use of the particular ability to be measured.

The cardinal principle of testing is that you should be testing what you want to test. You should therefore be very clear about what it is you want to test and try to ensure that you test only that. Tests should be valid. As far as possible your test should measure only what it is intended to test and not irrelevant capacities. The test writer must be clear about what an item is designed to measure and how performance on that item will be assessed before the candidates sit the test.

For example if you want to test for a knowledge of the present perfect tense in English you should make every effort to ensure that is the only thing your items are testing. In item 5 no indication is given of the criteria that will be used to assess success in the task and given each task will result in a different text type it is difficult to see how comparability between students doing different tasks is likely to emerge from such an item. The test writer must be clear about what an item is designed to measure and how performance on that item will be assessed before the candidates sit the test.

In achievement testing, tests should relate closely to both the content of what has been learned and also the way things have been taught. Students nurtured through the grammar translation approach are unlikely to perform well on tasks they have had no previous exposure to, such as spontaneous oral interaction. It may well be that to encourage teachers to adopt techniques that go with a new syllabus they need to be included in tests to increase the likelihood of their doing so. To the extent that tests are mirrors of the teaching there will be a positive washback effect from tests to teaching.

GENERAL CONCLUSIONS

(with reference to Bachman & Palmer, 2000)

- **Language testing should necessarily be related to language teaching and language use.** The score from a language test is an indicator of individual's language ability. It can be used to make certain decisions, especially in case of a "high-stakes" test. So *make sure* that performance on the test is related to language use in non-test situations – test tasks and situations should correlate with the language use tasks and situations. Characteristics of a test taker (his/her background knowledge, cognitive schemata, language ability) should also correspond to those of a language user.
- **Considerations of fairness should be built into testing.** Fairness is treated as being related to both the validity of a specific test as an index of certain ability and to the testing process as a whole. Those who are in test design and test administration are to be accountable for the way the test and its results are used. What is to be avoided is "over-nursing" the candidates: the latter should be treated

as independent and responsible individuals rather than helpless kids in constant need for assistance Any kind of a “halo-effect” in interpreting test results should also be a strong “taboo” in testing.

- **Test-takers (as well as stakeholders at large) should be provided with as complete information about the entire testing procedure as possible.** This will help to humanise the testing process in at least two possible ways – encouraging and enabling testees to perform at their highest level of ability on the one hand, and creating conditions for washback to follow in language teaching process, on the other hand.
- **The individual test qualities cannot be evaluated independently, but must be evaluated in terms of their combined effect on the overall usefulness of the test.**
- **Test usefulness should be an overriding consideration for quality control throughout the process of designing, developing and using a particular language test.**

REFERENCES

1. Alderson, Ch.J. (2005) *Assessing Reading*. Cambridge: Cambridge University Press.
2. Bachman, L.F. and A. Palmer (2000) *Language Testing in Practice*. Oxford: Oxford University Press.
3. Brown, H.D. (2001) *Teaching by principles: An interactive approach to language pedagogy*. White Plains, NJ: Addison, Wesley, Longman
4. Hughes, A. (2003) *Testing for Language Teachers*, 2nd ed. Cambridge: Cambridge University Press.
5. Weir, C. (1993) *Understanding and Developing Language Tests*. Prentice Hall Europe.

KEYS TO TASKS

TESTING TERMINOLOGY QUIZ (1) (Task format – “word associates”)	TESTING TERMINOLOGY QUIZ (2) (Task format – “matching”)
1 marking	1 marking
2 test user	2 marker
3 marker	3 scale
4 band	4 cut-off score
5 test construct	5 band
6 matching	6 scoring
7 specification	7 double marking
8 feedback	8 test user
9 scale	9 specification
10 cut-off score	10 test construct

TASK / ITEM TYPES QUIZ (1) (Task format – “matching”)

1	2	3	4	5	6	7	8	9	10	11	12	13	14
a	-	f	j	c	k	d	-	i	b	e	h	g	l

TASK/ITEM TYPES QUIZ (2)

- | | |
|---|---|
| 1 Note expansion | 17 Banked gap-filling (words) |
| 2 Translation | 18 Register transfer |
| 3 Word-ordering/sequencing/unscrambling | 19 Note taking |
| 4 Multiple matching (words to pictures) | 20 Word formation |
| 5 Open gap filling | 21 Gapped text (clauses) |
| 6 Gap-filling (paragraphs) | 22 Word associates |
| 7 Text ordering/sequencing | 23 Sentence transformation/paraphrasing |
| 8 Short Answer Questions | 24 Multiple choice |
| 9 Matching headings to paragraphs | 25 Dialogue/ discourse completion |
| 10 Proofreading | 26 Phonemic discrimination |
| 11 Summary | 27 Punctuation |
| 12 Dictation | 28 Multiple matching (questions to texts) |
| 13 Role play | 29 Table completion |
| 14 Matching banked words to definitions | 30 Essay |
| 15 True/False | |
| 16 Listing | |

O. Kvasova (Kyiv)
L. Hnarpov's'ka (Sumy)

ЦІЛІ ВПРОВАДЖЕННЯ НЕЗАЛЕЖНОГО ТЕСТУВАННЯ

(із доповіді міністра освіти і науки України 2 березня 2008 р.)

- забезпечення рівного і справедливого доступу до якісної освіти для всіх без винятку молодих людей;
- проведення загальнонаціонального моніторингу справжнього рівня нашої середньої освіти;
- ревізія усталених і сформованих упродовж багатьох років традицій, які сьогодні стали вже гальмом для розвитку освіти;
- усунення нічим не виправданого психологічного навантаження на школярів, які з одного й того самого предмета складають екзамени двічі поспіль – спочатку у школі, а потім у вищому навчальному закладі.

Ннаповс'ка, L. Language Testing: Known and Unknown (notes on the workshop)
[Text] / L. Ннаповс'ка, O. Kvasova // Іноземні мови. – 2008. – № 2. – С. 44-52.