

## Test tasks: Criteria

### Criterion 1: Objective vs subjective test tasks

The label "objective" or "subjective" is attached to a particular test task with reference to the method employed in its scoring. If it can be scored objectively (whether it is human- or machine-scored), then the task is said to be an objective task. On the other hand, if the rater/scorer needs to apply some degree of judgment to score the task, then the task is said to be subjective (with different degrees of subjectivity).

### Criterion 2: Direct vs indirect test tasks

The label "direct" or "indirect" is attached to a particular test task with reference to the way in which language for assessment is elicited from the students. For example, if the goal is to assess pronunciation of a given sound or set of sounds, a task which asked the candidate to pronounce words containing those sounds would be a direct task. One which asked the candidate to identify how a given sound is pronounced from a number of options (for example, in multiple choice format) would be an indirect task. A direct task for assessing student writing ability would be to ask them to write a text, etc.

### Criterion 3: Action required from the candidate to fulfill the task

- Test tasks which require candidates to select the correct answer are known as "selection tasks" (e.g. multiple-choice tasks).
- Test tasks which require candidates to transform a given item of language (whether a word, a phrase or even a sentence) are known as "transformation tasks" (e.g. rephrasing tasks).
- Test tasks which require candidates to complete a given item (usually with a given word or phrase) are known as "completion tasks" (e.g. fill-in-the-gap tasks).
- Test tasks which require candidates to re-order or re-arrange items (words within a sentence, or sentences within a text, for example) are known as "re-ordering/re-arrangement tasks".

## Course materials: Test tasks

### In detail...

#### 1) Objective vs subjective test tasks

This dichotomy opposes two different methods of scoring test results, and nothing else.

##### Objective scoring...

- **No judgement is required** of the scorer as regards the language sample(s) produced by the test-taker.
- That the scoring of a test is objective means that a testee will score **the same mark** no matter which examiner marks the task.
- It brings **reliability** to the scoring process –both rater and interrater reliability. In other words, the more objective the scoring, the greater agreement there will be between 2 scorers scoring the same paper (inter-rater reliability) or between 2 scorings at different occasions by the same scorer (intra-rater reliability).
- Since there is **only one correct answer or a limited set of correct answers**, these can be scored mechanically.
- Objective tests are not necessarily confined to particular lg skills or components. However, vocabulary tends to lend itself to objective methods of assessment, while writing or speaking are almost necessarily scored in a subjective manner.
- On the whole, objective tests **require far more preparation** than subjective tests. Thus, in an objective test, the tester spends a great deal of time constructing each item as carefully as possible, attempting to anticipate the various reactions of the testees at each stage. The effort, **however**, is rewarded in the **ease of the marking**.

##### Subjective scoring...

- The scorer is required to **make some kind of judgement** about the lg sample(s) produced by the test-taker.

E.g.: Marking a composition or an oral presentation.

- **Different degrees of subjectivity**: impressionistic scoring of a composition is more subjective than the scoring of short answers in response to questions on a reading passage.

## Course materials: Test tasks

### 2) Direct vs indirect test tasks

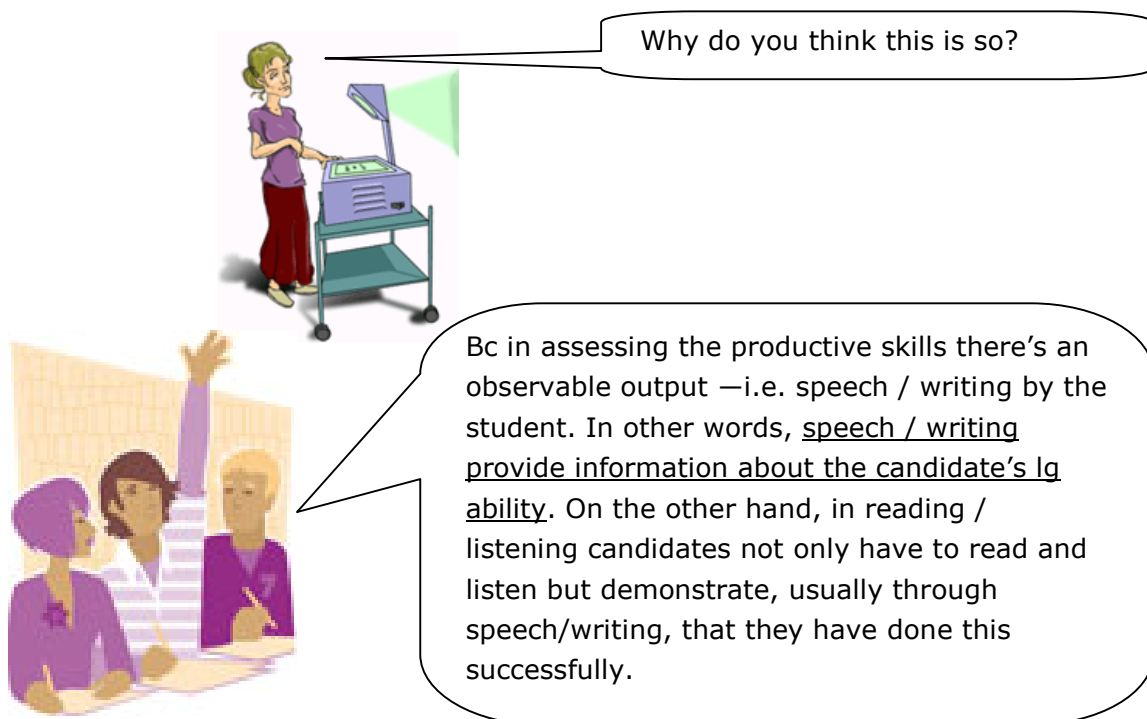
#### Direct testing...

- A test item is said to be direct when it requires the candidate to perform precisely the skill or language recognition / production task that we wish to assess.

E.g.1: To know how well students can write compositions, we ask them to write compositions.

E.g.2: To know how well students pronounce the L2, we ask them to speak.

- Direct testing is easier to carry out and therefore commonly associated with the productive skills (speaking & writing).



Why do you think this is so?

Bc in assessing the productive skills there's an observable output —i.e. speech / writing by the student. In other words, speech / writing provide information about the candidate's lg ability. On the other hand, in reading / listening candidates not only have to read and listen but demonstrate, usually through speech/writing, that they have done this successfully.

- Direct testing has a number of attractions:
  1. If we are clear about the abilities we want to assess, it is relatively easy to create the conditions which will elicit the behavior on which to base our judgements.

**Course materials: Test tasks**

2. In the case of the productive skills, assessing and interpreting student performance is quite straightforward.
3. Since practice for the test involves practice of the skills we want to foster, there is likely to be a **beneficial washback** effect.

Indirect testing...

- Indirect test items try to measure student knowledge & ability by getting at **what lies beneath their receptive & productive skills**.
- The design of procedures designed to tap into the enabling skills underpinning the macro skills results in indirect assessment devices of the skill in question.
- Thus, if we believe that grammatical knowledge contributes to writing ability, then a grammar test may be used as an indirect test of writing.

E.g.: One section of the TOEFL was developed as an indirect measure of writing ability. It contains items of the following kind where the candidate has to identify which of the underlined elements is erroneous or inappropriate in formal standard English.

At first the old woman seemed unwilling to accept anything that was offered her by my friend and I.

The ability to respond to such items has been shown to be related statistically to the ability to write compositions. However, the two identical are far from being identical. In other words, it's at least doubtful that this particular section can be used as a good measure of writing ability.

- Main problem: the relationship between performance on an indirect test and performance of the skill(s) in which we are usually more interested (productive skills) tends to be weak in strength and uncertain in nature.

E.g. Composition writing. We do not yet know enough about the component parts of composition writing as to predict in an accurate manner composition writing ability from scores on tests that measure the abilities that *we believe* underlie it.

**Course materials: Test tasks**

- Thus, to test general Ig ability, we should rely on a combination of indirect and direct testing and, if only on one of the two, this should be direct testing.
- If we are using only direct testing, then we have to make sure that we sample reasonably widely.

E.g. To test writing ability, we should collect at least two compositions, each calling on a different type of writing and on a different topic, so as to obtain more accurate estimates of the ability in question than would be obtained through indirect testing.



Can a very indirect test really provide valid assessment of the skill it is intended to measure?

- Does it mean that because someone is good at selecting the right answer on multiple-choice grammar items they are also effective writers?
- Does it mean that because someone can distinguish correctly between *ship* & *sheep* on a phonemic distinction test they will also be able to carry on a conversation effectively?

Summary: While direct test items try to resemble real language use as much as possible, indirect items try to find out about student language knowledge through more controlled items, such as m-c or grammar transformation items. These tend to be quicker to design and easier to score. They also produce greater scorer reliability.

## Types of indirect test tasks

### 1. Multiple choice items (MCIs)

The journalist was \_\_\_\_\_ by enemy fire as he tried to send a story by radio.

a wronged    b wounded    c injured    d damaged

- MCIs have been considered ideal test instruments for measuring students' knowledge of grammar & vocabulary, mainly bc they are **easy to score** and, with the use of computer technology, the answers can be read by machines, not people, with the consequent **elimination of scorer error**.
- **Problems with MCIs:**
  - MC tests are **difficult and time-consuming to construct**.
  - MC tests do not generally lend themselves to the testing of language as communication, mainly because choosing the right answer out of 4 or 5 possibilities has little to do with how language is used in real life, and neither do the processes involved in that selection. In real-life situations, appropriate responses to different stimuli are produced rather than selected from several options. However, as long as it is remembered that MCQs test knowledge of grammar, vocab, etc. rather than the ability to use language, MCQs can be useful in different teaching & testing situations.
- **Features of MCIs:**
  1. Number of alternatives. The ideal # of alternatives is 5. A larger #, e.g. 7, would obviously reduce the chance element, but it would be extremely difficult if not impossible to find as many as seven good options for each item. Actually, most classroom tests use 4 options, not 5, precisely bc of this difficulty.
  2. Areas to be measured & number of items to be included. Bf starting the construction of any given test, the test-developer must decide on a) the areas that the test is going to measure with MCQs and b) the number of items that are going to be included in the test.
    - The MC test must be long enough to provide evidence of the t-t's performance & short enough to be practicable.



**Course materials: Test tasks**

- An excessively long test is undesirable bc, apart from being more difficult to administer, it would cause mental strain & tension among the students taking it.
- The # of items will depend on the level of difficulty, the nature of the areas being tested, & the purpose of the test. Usually, the teacher's experience will determine the length of a test for classroom use.

**3. Context:**

- Both linguistic & situational context are essential in using lg. Thus, test-developers should make sure they don't build tests that consist entirely of a series of decontextualized items, since that could lead the students to thinking that lg is learnt & used free of any context.
- Isolated sentences in a MC test add to the artificiality of the test situation and create ambiguity and confusion on the part of the student.
- Awareness of the use of lg in an appropriate and meaningful way so essential in communication becomes irrelevant in a test made of decontextualized items. That is, all you are trying to teach your students in class becomes blurred by giving them decontextualized items in an exam.

- **Components of MCIs:**

a) The stem

The journalist was \_\_\_\_\_ by enemy fire as he tried to send a story by radio.

b) The responses (alternatives/options)

a wronged   b wounded   c injured   d damaged

One option is the answer, correct option or key, while the others are called distractors, bc their function is to distract weaker students (i.e. those who do not know the answer) from finding the correct option.

- **Principles of MC test construction:**

1. Each MC item should have only 1 answer. This answer must be ABSOLUTELY CORRECT unless the instruction says CHOOSE THE BEST OPTION or the like.

**Course materials: Test tasks**

2. Only 1 feature at a time should be tested, since it's less confusing for the t-t & helps reinforce a particular teaching point (the one that is measured).

Normally, nobody tests grammar & vocabulary at the same time. However, sometimes word order & sequence of tenses can be tested simultaneously. These are impure mc items:

**I never knew where \_\_\_\_\_.**

**a had the boys gone  
c have the boys gone**

**b the boys have gone  
d the boys had gone**

3. Each option should be grammatically correct when placed in the stem, except of course in the case of specific grammar test items.

**Someone who designs houses is a \_\_\_\_\_.**

**a designer**

**b builder**

**c architect**

**d plumber**

**John soon returned to \_\_\_\_\_.**

**a work**

**b the prison**

**c home**

**d school**



The 1<sup>st</sup> item can be corrected in at least 2 ways? How?

Stems ending in preps can also present some difficulties. In Item 2 (reading comprehension item), one option can be immediately ruled out. Which one?

4. All mc items should be at a level appropriate to the testee's proficiency level. The context should be at a lower level than the actual problem which the item is testing.

E.g.: A grammar test item should not contain other grammatical features as difficult as the area being tested.

E.g.: A vocabulary item should not contain more difficult semantic features in the stem than the area being tested.



**Course materials: Test tasks**

5. MCQs should be as brief and clear as possible (though it is desirable to provide short contexts for grammar items).

6. It is a good idea to put items in rough order of increasing difficulty. It's important to have a couple of simple items to lead in the t-ts, especially if they are unfamiliar with MC tests.

- **Characteristics & functions of item components (stem / options)**

**THE STEM**

Purpose: to present the problem clearly & concisely.

☛ The student should get from the stem a general idea of the problem and the answer required.

☛ The stem should not contain extraneous information or irrelevant clues which confuse the problem being tested. Unless students understand the problem being tested, there's no way of knowing whether or not they could have handled the problem correctly.

Form:

**a) an incomplete statement**

**b) a complete statement**

**c) a question**

He accused me of \_\_\_\_\_ lies.

**a speaking   b saying   c telling   d talking**

Everything we wanted was *to hand*.

**a under control                      b within reach**  
**c well cared for                      d being prepared for**

According to the writer, what did Tom immediately do?

**a He ran home.                                      b He met Bob.**  
**c He began to shout.                              d He phoned the police**

Contents:

**Course materials: Test tasks**

A MCQ should contain those words or phrases which would otherwise have to be repeated in each option.

The word 'astronauts' is used in the passage to refer to \_\_\_\_.

- a) travelers in an ocean liner.
- b) travelers in a space-ship.
- c) travelers in a submarine.
- d) travelers in a balloon.

The stem should be rewritten so that it reads:

The word 'astronauts' is used in the passage to refer to travelers in \_\_\_\_.

- a) an ocean liner.
- b) a space-ship.
- c) a submarine.
- d) a balloon.

The same principle applies to grammar items. The item below:

I enjoy \_\_\_\_ the children playing in the park.

- a) looking to
- b) looking about
- c) looking at
- d) looking on

should be rewritten like this:

I enjoy looking \_\_\_\_ the children playing in the park.

- a) to
- b) about
- c) at
- d) on

The first of these 2 items would be correct only if one of the errors made by the students in their free written work is the omission of the preposition after *look* (a common error), i.e. if it's clear they don't know that a preposition is necessary after this verb.

**THE CORRECT ANSWER**

- There should be no doubt as to the correct answer. Thus, each item should be checked by another person.
- The correct answer should be of approximately the same length as the distractors, especially in vocab. tests & tests of reading & listening comprehension. So, avoid tendency to make correct answer longer than distractors simply bc it is necessary to qualify a statement or word to make it absolutely correct. Here is an example of such a 'giveaway' item:



He began to *choke* while he was eating the fish.

- a) die
- b) cough and vomit
- c) grow very angry
- d) be unable to breathe because of something in the windpipe

### THE DISTRACTORS

- Each distractor should be reasonably attractive & plausible. It should appear right to the student who's unsure of the correct option.
- Items should be constructed in such a way that students get the correct answer by direct selection, not by elimination of obviously incorrect options.



The present tax reforms have benefited \_\_\_\_ poor.

- a that                      b the                      c a                      d an

- In general, distractors should be grammatically correct when standing by themselves. Otherwise, testees would be exposed to incorrect forms.
- Plausible distractors should be based on a) mistakes in the students' own written work, b) their answers in previous tests, c) the teacher's experience &, d) a contrastive analysis bt students' L1 and the L2.
- Distractors should not be too difficult nor demand higher proficiency than the correct option. If they do, they will only succeed in distracting the good student, who will think that they correct answer is too easy (and a trap).

### WRITING THE TEST

The testees will be required to perform any of the following tasks:

1. Write out the correct option in full in the blank.

He may not come, but we'll get ready in case he does.

- a will                      b does                      c is                      d may

(Not advisable, though, as I explained in class.)



## Course materials: Test tasks

2. Write only the letter of the correct option in the blank or in a box.

He may not come, but we'll get ready in case he \_\_\_\_\_.

a will            b does            c is            d may

B

3. Put a tick or a cross at the side of the correct option or in a separate box.

He may not come, but we'll get ready in case he \_\_\_\_\_.

a will            A.           

b does            B.           

c is            C.           

d may            D.           

4. Underline the correct option.

He may not come, but we'll get ready in case he \_\_\_\_\_.

a will            b does            c is            d may

5. Put a circle around the letter as the side of the correct option.

He may not come, but we'll get ready in case he \_\_\_\_\_.

a will            (b) does            c is            d may

As a way of summary, remember that in spite of all its attractions, m-c items have a series of limitations:

- M-c items test only recognition knowledge. E.g.: A m-c grammar test score may be a poor indicator of someone's ability to use grammatical structures. In other words, someone who can identify the correct response in a m-c item may not be able to produce the correct form when speaking or writing.
- Guessing may have a considerable but unknowable effect on test scores. To minimize the effect of guessing, we should definitely try to write four or five options for each item, and not just three.



**Course materials: Test tasks**

- M-c items severely restrict what can be tested bc they require distractors, and it may not always be possible to create them. E.g. the distinction bt present perfect and past simple in English. At the upper intermediate / advanced levels, it may not be possible to create distractors that really distract test-takers and, thus, even if you wanted to test this grammar aspect, you wouldn't be able to do it if you only used m-c items.
- It's very difficult to write good m-c items. When writing a m-c items, one needs to be careful
  - not to write 2 options that are correct;
  - not to forget to include one option that is correct;
  - not to give clues in the stem as to the correct option (e.g. not to use options that are much shorter or longer than the correct option, since this usually gives it away);
  - not to include ineffective distractors (i.e. distractors that do not distract).

This is such a time-consuming activity, that it is a good idea to keep an item bank from which items can be selected for any given test. Whether we decide to repeat items from test to test or not, that's up to the test-giver.

My experience: you can repeat a % of them from one test (e.g. June) to another (e.g. September) even though the students who take the second test have already taken the first one (which, however, is not always the case in our university system). From year to year, most students are usually new students, so one can decide to repeat more items. Obviously, whether one repeats items or not is also related to a very important factor: that the course contents and objectives have not changed from one test situation to the other.

- Backwash may be harmful. Needless to say, if a test that is very important to the students is m-c in nature, it's very likely that practice for the test may have a harmful effect on the learning and teaching processes. Practice at m-c items is not the best way to improve students command of a language, especially when a large part of that practice is devoted to improve one's guessing ability, not just one's ability in the L2.
- Cheating may be facilitated. There's no need to say that the fact that the answers are so simple (a, b, c...) makes them easy to communicate to other students in a non-verbal way. How can you avoid this? By having Option 1 and Option 2 of the same test, with the items arranged in a different order.

This all seems to suggest not that we ought not to use them, but that in any given test we should not just use indirect testing / objective marking items, but direct and subjective ones as well.

## 2. The cloze procedure

- In its purest form, a cloze consists of the deletion of every nth word in a text (somewhere bt every fifth or tenth word). Because the procedure is random, it avoids test designer failings.

Example of cloze fragment:

They sat on a bench attached 1 \_\_\_\_\_ a picnic table. Below them they 2 \_\_\_\_\_ see the river gurgling between overgrown 3 \_\_\_\_\_. The sky was diamond blue, with 4 \_\_\_\_\_ white clouds dancing in the freshening 5 \_\_\_\_\_. They could hear the call of 6 \_\_\_\_\_ and the buzzing of countless insects. 7 \_\_\_\_\_ were completely alone.

- Because of the randomness of the deleted words, anything may be tested: grammar, collocation, fixed phrases, reading comprehension... Which makes it, at least on the face of it, the perfect testing instrument.

### PROBLEMS WITH THE CLOZE-PROCEDURE

However, ...

- the score obtained by the students depends on the particular words that have been deleted, rather than on their general knowledge of the lg;
- some items are more difficult to supply than others;
- in some cases, there may be several possible answers;

In spite of these reliability problems, supplying the correct word for a blank does imply...

- an understanding of context &
- a knowledge of that word and how it functions,

which makes the cloze technique very useful in lg tests.

### CLOZE OR FILL-IN-THE-BLANKS/GAPS?

Cloze tests may look similar to completion or blank-filling tests, but they are different:

- In blank or gap filling tests, the word for deletion is selected subjectively (consisting largely of structural words in grammar tests & key content words in vocabulary tests), whereas in cloze tests the words are deleted systematically.

- Gap filling items belong to the discrete-point approach to language testing whereas the cloze procedure is known to be one of the most representative techniques of the integrative approach to language testing.

### CLOZE CONSTRUCTION

Once the text has been chosen, the construction of the cloze is purely mechanical: every "nth" word is deleted.

- Deletion interval: commonly bt every 5<sup>th</sup> & every 10<sup>th</sup> word. BUT, if every 7<sup>th</sup> word has been deleted in the first few sentences, that is the interval that should be used for the rest of the text.

- 5<sup>th</sup>, 6<sup>th</sup>, and 7<sup>th</sup> word intervals are the preferred, mainly because a shorter interval would make it very hard for the student to just understand the text, since there would not be enough context. If, on the other hand, every 10<sup>th</sup> or 12<sup>th</sup> word is deleted, it would be necessary to have a very long text. Examples:

E.g.1: To make 40 deletions every 7<sup>th</sup> word ◀ a 280 to 300-word text would be needed.

E.g.2: To make 40 deletions every 12<sup>th</sup> word ◀ 480 to 500-word text would be needed.

### CLOZE SCORING METHOD

1. Exact word method: students get credit for a correct answer if and only if the word they write in any given blank is the exact word deleted from the original text. This is approach is quick and, therefore, very practical, and also highly reliable.

## Course materials: Test tasks

PROBLEM: the exact word scoring method may be too rigid –i.e. it does not reward creativity on the part of the test-taker.

2. Acceptable word method: any response that (a) is grammatically correct and (b) makes good sense in the context is given full credit as an acceptable answer. This method may promote positive washback, since it could encourage learners to use their pragmatic expectancy grammars creatively.

PROBLEM 1: it may slow down the scoring process, specially if you have a large # of students.

PROBLEM 2: it could affect scoring reliability if scorers don't agree about the acceptability of some of the words supplied by the students before the actual scoring takes place.

### LEVEL OF CLOZE TEXT DIFFICULTY

The level of difficulty of the text is very important: if the text is already difficult to read without blanks, imagine how difficult it would be once the blanks are inserted!

The difficulty level of the text is affected by as many as the following variables:

- ☛ text length;
- ☛ amount of time allowed to complete the task;
- ☛ learner familiarity with vocab & syntax of the passage;
- ☛ length & complexity of the sentences in the passage;
- ☛ learner familiarity with topic & with discourse genre of text (content & formal schemata); and
- ☛ blank interval (every 5<sup>th</sup> word vs every 9<sup>th</sup> word, for instance).

### MUTILATION

Once you've selected or written the text, you have to decide on the interval at which you will be eliminating or *mutilating* words. Basically, there are 2 ways to mutilate a text:

- a) rational deletion (or selected deletion): test developer deletes words on the basis of some rational decision. E.g.: to test students' knowledge of verb tenses, delete only verbs. (Some writers say this is not really a cloze test, but a completion test).





**Course materials: Test tasks**

- b) fixed ratio or nth word deletion: regardless of its part of speech or the semantic load it bears within the text, every nth word is omitted. [b) more difficult than a) for the student]

**3. C-Test**

This is a variation on the cloze test, in which the students read a brief paragraph in the L2. The first two sentences are left intact. There\_ \_ \_ \_ , every ot\_ \_ \_ word i\_ \_ printed int\_ \_ \_ , but f\_ \_ each alte\_ \_ \_ \_ \_ word, on\_ \_ the fi\_ \_ \_ half o\_ the wo\_ \_ is wri\_ \_ \_ , and t\_ \_ second ha\_ \_ is indi\_ \_ \_ \_ \_ by a bl\_ \_ \_ space repres\_ \_ \_ \_ \_ each let\_ \_ \_ . T\_ \_ students' abi\_ \_ \_ \_ \_ to fi\_ \_ in t\_ \_ blank spa\_ \_ is tho\_ \_ \_ \_ to b\_ a mea\_ \_ \_ \_ of th\_ \_ \_ language profi\_ \_ \_ \_ \_ .



***This is a variation on the cloze test, in which the students read a brief paragraph in the L2. The first two sentences are left intact. Thereafter, every other word is printed intact, but for each alternate word, only the first half of the word is written, and the second half is indicated by a blank space representing each letter. The students' ability to fill in the blank space is thought to be a measure of their language proficiency.***

• This approach to text mutilation is called the rule of two -i.e. starting with the 2<sup>nd</sup> sentence of the text, the 2<sup>nd</sup> half of every 2<sup>nd</sup> word is deleted. In words having an odd number of letters, there are more blanks given than letters (e.g., thought is represented as tho\_ \_ \_ \_), but this pattern can be altered to suit the needs of particular groups of students.

• It's been proved that C-tests can also be excellent teaching devices (apart from testing techniques) because they provoke creative reasoning among the students, especially if they do it in pairs or groups: student lg use focuses on the lg as

**Course materials: Test tasks**

content (“that’s not enough letters” for example, they might say, or “we need an adjective here”, etc.).

The scoring method for C-tests is very precise. If a letter is wrong, or if the test-taker writes more or fewer letters than are called for, the item is marked as incorrect.

A word of caveat: the successful completion of C-tests does not only depend on L2 ability, but on background knowledge (or lack of it). In other words, knowing about the topic of the text also helps to decide how the words in it can be completed. This is related to the schema theory, according to which content schemata as well as formal schemata influence our ability to process texts.

Exercise: Try to complete the following C-test. When doing it, try to be aware of your own mental work: by what psycholinguistic processes do you make decisions about what the missing letters should be?

Don’t panic! Cultural content makes this text quite challenging!

An architectural feature that is basically a mass of fine sculpture is the Sinhalese staircase. It includes an intricately decorated flight of steps flanked by ornate balustrades with a frontal stand formed by a pair of guardstones.

(1) The elabo\_\_\_\_\_ moonstone a\_ its ba\_\_ is i\_ itself /\_ / distinctive ele\_\_\_\_\_ of anc\_\_\_\_\_ sculpture i\_ the isl\_\_\_\_\_. (2) These se\_\_-circular sto\_\_\_\_ acquired increa\_\_\_\_\_ complex ba\_\_ of decor\_\_\_\_\_ over t\_\_ years. (3) Th\_\_ range fr\_\_ the abst\_\_\_\_\_—tongues o\_ fire a\_\_ bands o\_ creeper vi\_\_\_\_—to symb\_\_\_\_\_ interpretations o\_ the fo\_\_ perils o\_ life. (4) T\_\_ latter con\_\_\_\_\_ of t\_\_ elephant, /\_ / symbol o\_ birth; t\_\_ bull, indic\_\_\_\_\_ of de\_\_\_\_; the li\_\_ representing dis\_\_\_\_\_; and t\_\_ horse, /\_ / symbol o\_ death. (5) So\_\_ also ha\_\_ a ba\_\_ of ge\_\_\_\_, which repre\_\_\_\_\_ the disti\_\_\_\_\_ between go\_\_ and ev\_\_. (6) At t\_\_ heart o\_ many moons\_\_\_\_\_ is /\_ / lotus pe\_\_\_\_. (7) Buddhists reg\_\_\_\_ the lo\_\_\_\_ as /\_ / sacred flo\_\_\_\_, and i\_ the moon\_\_\_\_\_ it repre\_\_\_\_\_ the cen\_\_\_\_ crowning se\_\_ of suc\_\_\_\_\_.

The total effect, then, of mounting the stairs begins with the devotee at the lotus, his back to the world, preparing to ascend to the shrine of wisdom and insight.



(1) The elaborate moonstone at its base is in itself /a/ distinctive element of ancient sculpture in the island. (2) These semi-circular stones acquired increasingly complex bands of decoration over the years. (3) They range from the abstract—tongues of fire and bands of creeper vines—to symbolic interpretations of the four perils of life. (4) The latter consist of the elephant, /a/ symbol of birth; the bull, indicative of decay; the lion representing disease; and the horse, /a/ symbol of death. (5) Some also have a band of geese, which represents the distinction between good and evil. (6) At the heart of many moonstones is /a/ lotus petal. (7) Buddhists regard the lotus as /a/ sacred flower, and in the moonstone it represents the center crowning seat of success.

#### POSITIVE ASPECTS OF C-TESTS

- C-tests are not only testing techniques, but they can also be used as very useful teaching techniques, mainly because many of the grammatical morphemes load at the end of words in English (e.g. plurals, tense markers, aspect markers, etc.).
- C-tests seem to foster discourse-level processing. For instance, you may have decided that the second sentence begins with “These semicircular stones acquired...” because you realized that the topic and grammatical

**Course materials: Test tasks**

subject of the preceding sentence was *moonstones*. Likewise, many people initially complete the 4<sup>th</sup> sentence with “The latter consists of the elephant, a symbol of birth; the bull, indicative of *death*, ...” (perhaps in association with *birth* in the previous phrase). But then a little further in the same sentence we come to the phrase, “and the horse, a symbol of *death*.” At this point we realize it would be unlikely that both the bull and the horse would be symbols of death, given the listing structure of the paragraph (and this insight is motivated in part by our formal schemata). It is a given that the horse symbolizes death, since the word *death* is printed in its entirety in that clause. So here we are forced to reconsider our options for filling the blanks in the clause, “the bull, indicative of de\_ \_ \_.”

