# Testing reading comprehension in an ESP context: some tips

**Nadia MENEZLA**
*Department of English Faculty of Humanities Djilali Liabes University of Sidi Bel Abbes, Algeria*
***Mail:menezlanomnadie@yahoo.fr***

***Abstract–****Researchers assure that in high education, and in other programs, which use teaching subjects written in English, reading becomes the pioneer among all the other skills of the English language. Accepting the fact that reading is of great importance for all students, it is attached high importance for ESP learners. But traditional methods in testing it are problematic. The purpose of this paper is to provide some practical suggestions for testing reading comprehension in English for Specific Purposes (ESP). In the light of some common-sense criteria of testing, it begins with an outline of these criteria and ends up with some suggested testing procedures.*

***Key-words:*** *ESP, reading comprehension, testing reading comprehension.*

## Introduction

The problem is to develop easily-constructed tests which both valid and reliable. To be valid they must test skills which correlate well with the skills students need (or will need) to use. To be reliable, they must produce a result which is repeatable -a similar result if corrected by

another examiner; a similar result if the student takes another test for the same type and length. In other words, we must be sure we are testing the right skills (validity) and providing a true measure (reliability). Unlike listening comprehension or writing, almost all tests of reading comprehension in ESP have serious weaknesses in either reliability or validity.

The literature on the subject is frustrating. Reading comprehension testing is generally assumed to mean providing a text and a series of multiple-choice questions which follow it. The deficiencies of MC questions are often pointed out, but useful alternatives are usually not provided. What do we want these tests for?

## Background

The term testing is famous for its various uses depending on the author. For the purposes of this short paper it will be a synonym of formal assessment (Harris & McCann, 1994). Tests can be for a variety of purposes but the most common which can be stated here is to find out to what extent the student the student has learned. Now the student could have learned, or part-learned a) some skills, or b) some facts, or c) some ideas, or a combination. In this paper we shall concentrate on a), while not forgetting that ESP should also involve b) and c). Wondering what a) involves, leads us to ask what the teaching objectives – what skills do we want our students to learn?

Comprehension can be seen as a scale, with a 100% comprehension at one end, and no comprehension on the other. If we divide this skill into three, we end up with three levels of comprehension:

- General comprehension – knowing what the topic of the text, and having a rough idea of what the author says about the topic.
- Main points comprehension- general comprehension + understanding the main ideas, without grasping details.
- Detailed comprehension- main points comprehension + understanding all or most of less important details.

It seems that ESP reading comprehension courses have generally concentrated on level three- detailed comprehension, which is the bottom of the list but the top of a pyramid (Shabaan, 2005). We notice

that in our native language we often aim at level one or at level two, not just at level three respectably and reasonably, when deciding what we read. And for purpose of *study* it is highly advisable to cultivate the ability to extract basic ideas (levels one and two) first, and then return to a text for more detailed comprehension if merited. The above should be recognizable as an argument for teaching reading comprehension an all three levels, not just on the level of detailed comprehension. Being so, it seems logical to aim out testing at all three levels, not just level three.

**Some commonly used test procedure**

*Open-ended questions:* at first sight, the procedure of testing WH questions, which requires extended answers, seems perfectly reasonable. But sometimes the problem is not that of validity perhaps. The difficulty is sometimes the short scoring- what exactly will account as a short answer? In practice it is impossible to predict all possible right answers; the scoring is therefore a matter of discretion and subjectivity, so that different examiners will score the test paper differently. Furthermore, the kind of question which is more objective looking, such as one beginning who…? Or "when..?" (and therefore, more reliable) is inherently artificial.

*Summary-type tests:* these are probably less artificial, in that they require the reader to reduce a mass of language to its main points, which is quite a likely task for real life. But they are again inherently unreliable. Besides, summary writing is a rather specialized task and needs special training- it would be possible to comprehend well, even at level three, and not know how to write a suitable summary.

*Multiple choice questions:* it was in response to this lack of reliability that MC format evolved. The problems here are many, and have been well documented. Briefly, there is a real danger that:

- the options give the answers away;
- one can answer without reading the text;
- the incorrect options (distracters) do not distract;
- there is no correct answer;
- there too many correct answers;

- choosing the right answer becomes quite difficult.

Even if all these traps are avoided (and it is very hard to avoid them all, even with great patience, skill, knowledge, help, etc, the language of the text and of the questions, like all language, being liable to personal interpretation), there remains a serious problem of validity- the questions, it seems, will tend to ask about details, not about the main points. For every question which asks about the main ideas, topic, or theme, there will be some questions which ask about relatively minor details (Basturkmen, 2006). And this is not surprising, since the MC format demands that one should give the student information, in the lead-in to the question, or in options themselves. The test-writer never wants to give more information away, and so tends to ask about the details where the necessarily-given information will not affect the student's knowledge of the rest of the text. MC questions will therefore naturally tend to towards detail.

*True-false questions:* The same objection as with MC, the problem is again one of validity rather than reliability. Actually, TF is just a variant of MC; only allowing two options, it gives less away to the student but also informs the examiner less-after all, in MC the choice of *which* incorrect option can tell the examiner something, whereas TF answers only either right or wrong, and may well be so by chance(Harris & McCann,1994).

*Close testing:* great claims have been made for close tests, wherein every seventh is deleted, the object being to restore the message to a mutilated text bi filling in the blanks. At first sight this appears to be a thoroughly artificial procedure, but it can be argued that this 'restoration' is what the reader does naturally anyway in real situations, where 'noise' interferes with 'reception'- poor quality printing, typing mistake, and so on (Hughes & Schumaker,1991).

There are numbers of modifications to the close procedure, and as many different scoring systems. But as far as we are concerned, in the reading comprehension ESP class, and as the close procedure involves *writing,* it is requiring an additional skill. Further, the student who fails to score on one blank because he puts, for example, the wrong

preposition, loses the point (in most scoring systems) despite the fact that he recognized correctly that what was needed was some kind of preposition.

**Some further considerations**

How many items? For a test to be informative to the examiner, it must include many items, in other words, many possibilities of error. A score of 1 out of 2 is less informative than 10 out of 20- the fewer the items, the higher the possibility of a high score (or a low one or an average one) being because of chance. On the other hand, a reasonable time limit precludes asking a lot of detailed questions on several long texts.

**A summarizing table**

| ***Procedure*** | ***Level of comprehension*** | ***problems*** | ***advantages*** |
|---|---|---|---|
| Open-ended questions | Tends toward details | Unreliable in scoring, artificial | Easy to construct |
| Summary-type | Main points | Unreliable in scoring | Easy to construct |
| Multimle-choice | Tends to details | Very hard to construct, tends to rely on fine distinctions | Easy to score |
| True-false | Tends to details | Hard to construct, chance influences results | Easy to score |
| Cloze | Details | Requires writing, requires detailed knowledge of grammar | Easy to construct, fairly easy to score, high number of items per test |

**Suggested techniques**

Problems have been shown to crop up with five testing procedures. The three procedures also have their associated problems- nothing in

testing is as simple as it seems at first- but are put forward for consideration, and can all be corrected mechanically.

**1-For detailed comprehension:**

*Cloze with different word-class MC options:* This procedure entails deleting every seventh word, as in the classical cloze procedure, but numbering the spaces and offering four or five alternatives to fill in the space.

*The alternatives belong to different word-classes:* The task is therefore one of recognizing what sort of a word goes in the space. The procedure is still weighted towards level of comprehension-detailed comprehension- as it requires the student to follow everything carefully, word by word. It is however, easy to construct and to score, and has at least one major advantage over the classical cloze procedure: it does not require writing. Preliminary investigations suggest that it is easier than other testing techniques, so the time factor becomes more important.

**2- For main points comprehension:**

*Paragraph and sentence reordering:* This technique entails putting back into correct order a text whose paragraphs have been jumbled, or whose sentences have been jumbled. Here is an example, with jumbled paragraphs. The instructions ask the student to note the correct order of paragraphs, using the numbers in the left-hand margin (appendix 1). This technique is easy to correct if well constructed, but does not necessarily yield many items. It is probably best suited to the functions or narratives, where it is a clear correct order.

*Finding the extra irrelevant sentence:* This procedure is easy to construct and to correct. The examiner inserts extra, irrelevant sentence in a text. The student must indicate which sentences are irrelevant. The sentences are numbered in the right hand margin (appendix2). This procedure has the disadvantage that it is artificial- in real life texts do not come interspersed with irrelevant extra sentences. However, a case can be made out that the skills required to find the irrelevant sentences

involve the comprehension skills of appreciating the main points. Also, students like the procedure, and it seems to produce satisfactory results.

**Conclusion**

Being aware of the great importance of reading for ESP learners, we have to recognize that testing it is difficult and may often upset people, particularly if the scoring system is unfair, or if the procedure is hard understand, or if the skills involved do not seem relevant or necessary. That is why, as teachers we should be interested in hearing from anyone who has tried out these or other similar procedures.

## References

Baker, L., Dreher, J. & Guthrie, J. T. (2000). *Engaging young readers: Promoting achievement and motivation.* New York: Guilford Press.

Basturkmen, H. (2006). *Ideas and options in English for specific purposes*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Belcher, D. (2006). English for specific purposes: teaching to perceived needs and imagined futures in worlds of work, study, and everyday life. *TESOL Quarterly.* 40(1) pp. 133-156.

Blanton, W. E., & Wood, K. D (1984). Direct instruction in reading comprehension test-taking skill. *Reading World*, 24, 10-19.

Green, D. S, & Bushell, D. (1976). The effects of an instructional test-taking unit on achievement test scores. *Journal of Educational Research*, 70(2), 67-71.

Guthrie, J. T., & Alao, S. (1997). Designing contexts to increase motivations for reading. *Educational Psychologist*, 32, 95-107.

Haladyna, T. M., Nolen, S. B., &Hass, N. S. (1990). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2-7.

Hughes, C. A., & Schumaker, J. B. (1991). Test-taking strategy instruction for adolescents with learning disabilities. *Exceptionality*, 24, 205-221.

Harris, M. & P. McCann (1994). *Assessment*. Oxford: Heinemann.

Master, P. (2005). Research in English for specific purposes. In E. Hinkel (Ed.), *Handbook of research n second language teaching and research.* (pp. 99–115). Mahwah, NJ: Lawrence Erlbaum.

Millman, J., & Bishop, C. H. (1965). An analysis of test-wiseness. *Educational and Psychological Measurement,* 25(3), 707-717.
Shabaan, K. (2005). An ESP Course for employees at the American University of Beirut. *ESP World.* 4, 2- 10.

# Appendix 1
# Paragraph and sentence reordering

1- In other studies, both birthdays and bachelorhood have been implicated as risk factors. Statistics reveal that heart attack rates increase significantly for both females and males in the few days immediately preceding and following their birthdays. And unmarried men are more at risk for heart attacks than their married counterparts. Though stress is thought to be linked in some way to all of the aforementioned risk factors, intense research continues in the hope of further comprehending why and how heart failure is triggered.

2- Cardiologists first observed this morning phenomenon in the mid-1980, and have since discovered a number of possible causes. An early morning rise in blood pressure, heart rate, and concentration of heart stimulating hormones, plus a reduction of blood flow to the heart, may all contribute to the higher incidence of heart

attacks between the hours of 8:00 a.m. and 10:00 a.m.

3- But according to new studies, the list of risk factors may be significantly longer and quite surprising. Heart failure, for example, appears to have seasonal and temporal patterns. A higher percentage of heart attacks occur in cold weather, and more people experience heart failure on Monday than on any other day of the week. In addition, people are more susceptible to heart attacks in the first few hours after waking.

4- As heart disease continues to be the number-one killer in the United States, researchers have become increasingly interested in identifying the potential risk factors that trigger heart attacks. High-fat diets and "life in the fast lane" have long been known to contribute to the high incidence of heart failure.

## Appendix 2

# Finding the extra irrelevant sentence

Not much later, perhaps 2.5 million years ago, the bipeds were
1
making stone tools and hunting animals for food. It seems obvious
2
that his action in trying to take part in strenuous walking or
running was a mistake in view of the publicity (mostly unfavorable)
that this collapse and racing heart-beat aroused. By about
3
two million years ago hominid craniums with a larger capacity
appeared. By 1.5 million years ago Homo Erectus was on the scene,
4
the brains had doubled in size and the stone tools now included
bifaces, tools that have been flaked on both sides. He was in
5
fact known throughout the land as a competent carpenter, but
nowadays his fame rest on the fact that he was the father of the
man who has been acclaimed as our savior. These bifaces belong
6
to the core-tool industry known as the Acheulian. (The characteristic
St. – Acheul) 7
curiously, France is one of the Eurapean
8
countries where a knowledge of the native tongue seems to be a
prerequisite for an enjoyable stay, as great reluctance is often
expressed by the French when they are expected to speak someone
else's language. From about two million to one million years ago
9
another kind of early biped was also present. Its robust anatomy
10
Identifies it as avseparate specie of Australopihecus. It is
11

readily distinguishible from the less robust biped by its massive
jaw and molar teeth that are very large compared with the incisors.
two or three million bipeds without teeth advancing across the
12
land in an angry search for a dentist would certainly be diconcertin
for all other humans who stood In their path.
this summary of the huminid fossil record is undoubtedly
13
overimplifie, but I think the evidene supports th mai outline.
what difficulties there are arise from the fragmentary nature of
14
many of the fossils. That is because they are known to have been
15
deliberatlly smashed by early hominids who wanted to make
archaeology
more difficult for their remote descendants. For example,
16
johnson has found one skeleton in the Afar region complete enough
to allow reconstruction of that hominid's general proportions. The
17
reconstruction shows that it had relatively long arms, a fact that
could not be determined from thre hundreds of previously discovered
fragmentary remains of Australopithecus.