UNIVERSIDAD TECNOLÓGICA NACIONAL

INSTITUTO NACIONAL SUPERIOR DEL PROFESORADO TÉCNICO

En convenio académico con la Facultad Regional Villa María

LICENCIATURA EN LENGUA INGLESA

Tesis de Licenciatura

ASSESSING VOCABULARY: DOES TEST FORMAT INFLUENCE TEST RESULTS?

Tesista

PROFESORA MARIANA VANESA CASTINO

Director de la Tesis

MAGÍSTER MARIANO QUINTERNO

2011

UNIVERSIDAD TECNOLÓGICA NACIONAL

LICENCIATURA EN LENGUA INGLESA

Dissertation

ASSESSING VOCABULARY: DOES TEST FORMAT INFLUENCE TEST RESULTS?

Candidate

PROFESORA MARIANA VANESA CASTINO

Tutor

MAGÍSTER MARIANO QUINTERNO

2011

Dedication

To the memory of my father, whose love and wise words

I will always keep in my heart.

Acknowledgements

I would like to express my sincerest gratitude to my tutor, Magíster Mariano Quinterno for his invaluable guidance through the process of writing this dissertation. It was thanks to his knowledge, expertise, and his uninterested dedication that I was able to achieve such a goal in my career.

I am immensely grateful to Doctor Omar Villarreal for his precious help whenever I needed it. I must also thank Magíster Marina Kirac and all the teachers in the Licenciatura en Lengua Inglesa at Universidad Tecnológica Nacional, who endowed me with their enlightenment and transcendent teaching.

I am also indebted to María Eugenia Alvarez, who has always supported and encouraged me with her understanding and friendly counselling.

I am particularly thankful to the heads of the schools of English in Florida, who allowed me to carry out the present study in their institutions; and to their teachers and students, who resigned time of their classes to make their contribution to this investigation.

I would like to express my love for my beloved ones, my husband, my two daughters Tiziana and Julieta, and my mother, who are the pillars of my life. They have been endlessly supportive and considerate and generously helped me overcome all the difficult instances that I encountered during this process.

Last but not least, my heartfelt thanks should go to my father, who generously waited for me to finish my research work before he departed from this World.

Abstract

Assessment has become a relevant part of the teaching and learning process in the past few decades. However, much less importance has been given to the assessment of vocabulary than to other components of the language, such as grammar. As a consequence of this, little attention has been paid to the influence that the methodology used in a vocabulary test can actually have on the results obtained. This study is aimed to explore to what extent test format influences test results in the assessment of vocabulary. For this purpose, First Certificate in English (FCE) students were given three vocabulary tests and an interview was also given to those students who showed some divergence among the results of the three tests. The results indicate that test format can greatly influence the results obtained in vocabulary tests. The findings also suggest that the successive task administrations, evaluating the same key words, do not lead to a better test performance.

Key words: assessment, tests, tasks, vocabulary, test format, test results

Resumen

La evaluación se ha convertido en una parte importante del proceso de enseñanza y de aprendizaje en las últimas décadas. Sin embargo, mucha menos importancia se le ha dado a la evaluación de vocabulario que a otros componentes del idioma, como por ejemplo a la gramática. Como consecuencia de esto, se le ha brindado poca atención a la influencia que la metodología usada para evaluar vocabulario puede tener en los resultados que se desprenden de dichas evaluciones. Este estudio se propone investigar hasta qué punto el formato de las evaluaciones de vocabulario influye en los resultados que se desprenden de las mismas. Con este propósito, tres evaluaciones de vocabulario se tomaron a estudiantes preparando el examen internacional "First Certificate in English" y se entrevistó a aquellos alumnos cuyos resultados obtenidos en las evaluaciones mostraron alguna divergencia. Los resultados indican que el formato de las evaluaciones de vocabulario puede tener gran influencia sobre los resultados que surgen de dichas evaluaciones. También se comprobó que la administración de ejercitaciones que evalúan las mismas palabras en forma consecutiva no hace que el alumno mejore el desempeño en dichas evaluaciones.

Palabras clave: evaluación, pruebas, ejercitaciones, vocabulario, formato de prueba, resultados de la prueba.

Table of Contents

CHAPTER I

Introduction

Assessment has nearly gained as much relevance as teaching and learning in the field of English Language Teaching. Even though some teachers still see assessment as something that occurs separately from teaching and learning, many others regard it as an integral part of the teaching and learning process. Thus, such relevant role of assessment in ELT may result from considering it as "one of the most valuable sources of information about what is happening in the learning environment" (Harris and McCann, 1994, p.2).

However, the test, the practical activity used to assess students' ability, skill, or knowledge, does not always show what it purports to measure. In the case of vocabulary tests, it sometimes becomes difficult for the test designer to design a test that can measure vocabulary knowledge as a separate component of language ability. Thus, this difficulty can lead to the inclusion of a construct, that is, some knowledge or ability, that was not intended to be measured in that particular test. Such being the case, the results of the test may appear to be different from what was expected. Therefore, the knowledge or ability to be measured becomes an important source of influence on test scores. Another source of great importance that influences test scores comes to be the testing

task or the method whereby the test takers' knowledge will be measured (Read, 2000).

According to these assumptions, a relevant question arises: *does test format influence test results?* Thus, the present study is informed by the following research question: *to what extent does test format influence test results in the assessment of vocabulary?*

As derived from the research question, the following hypotheses will be examined:

1. Test performance is affected by the test format used for the assessment of vocabulary.

2. Test performance is not affected by the test format used for the assessment of vocabulary.

3. Test performance improves with the successive administration of the same or similar tests for the assessment of vocabulary.

CHAPTER II

Literature Review

1 Assessment

1.1 Definition

According to the *Longman Dictionary of Contemporary English* (1995), *assessment* can be defined, in general terms, as a process in which someone passes judgment on a person or a situation. Shepard (2002) defines *assessment* as "the process of collecting data to measure the knowledge or performance of a student or group" (p. 2534). In the field of education, the term *assessment* can be ambiguously associated with other terms as *evaluation* and *tests*, which are similar in meaning and share some characteristics. *Assessment* can be considered to be the process whereby a teacher can measure students' performance and learning process, whereas *evaluation* can be used more widely to measure all the aspects that are influentially related to learning and teaching such as materials, syllabus, course, methodology, teacher performance and assessment, itself. Tests, on the other hand, enable the teacher to measure progress in a more individualised way and are only part of the assessment process (Harris and McCann, 1994). "The word *assessment* carries with it the idea of a broader and more comprehensive evaluation of

student performance than a single test" (Shepard, 2002, p. 2534). According to Shepard (2002), "in an age when testing is controversial, *assessment* has become the preferred term because of its connotation of breadth and thoroughness" (p. 2534).

1.2 The influence of testing on teaching and learning

"The influence of tests on teaching and learning is called the *washback effect*" (Harris and McCann, 1994, p. 27). Harris and McCann state that if students have to take a test at the end of the course, this will affect the syllabus. They also point out that a test can have a positive or a negative influence on teaching and learning depending on whether it contains authentic examples of the type of tasks which the learners will need to perform in the future or artificial tasks not linked to real future needs. When tests contain artificial tasks, Harris and McCann argue that "teaching methods will probably reflect these tasks and the learning process could end up revolving around what we might term 'exam practice'" (p. 27).

"Assessment is generally seen as something done to students by teachers" (Harris and McCann, 1994, p. 2). Students may feel panic and confusion when, from time to time, they are asked to take a test that has to be passed so as not to be considered a failure (Harris and McCann, 1994). Evidently, some

assessment procedures have contributed to this view. The feedback that the students receive from a written or oral test is generally expressed only by a mark, thus classifying the students but not providing the information that the students are supposed to receive when they are evaluated. This way, students lack important information about their progress or about their specific problems that need to be overcome. However, students sometimes receive this proper feedback about their performance and progress in an end-of-the-year test but this information comes much too late to be formative; thus assessment takes place after learning has finished, rather than during the learning process (Harris and McCann, 1994).

Harris and McCann (1994) also argue that another reason for negative attitudes towards assessment among students is that a test is generally meant to highlight students' mistakes, "...it tries to catch students out..." (p. 3), instead of giving students the opportunity to show what they have learnt. On the contrary, Lado (1961) claimed that a test should be designed in such a way that the students' weaknesses can be detected. He argued that teachers should select as the content of a test what is to be learned and should eliminate from it what is easy to master, such as knowledge from previous language training. In his words, "We say specifically that testing the problems is testing the language" (p. 20).

Another bad washback effect of assessment can be usually produced by the dichotomy between what is given and worked on in the classroom and what is actually tested. Harris and McCann (1994) attribute this drawback to the fact that when teachers have to evaluate students, they generally concentrate on only one part of what has happened in the classroom during the teaching and learning process. Even though most English teachers apply the Communicative Approach in their classes, many of them still give their students tests whose content is entirely grammatical. In reference to this matter, Skehan (1991) states that:

> From within language testing we have, unfortunately, a tendency to want to believe in old and all-embracing models. One example of this would be a belief in language consisting essentially of grammar, and which assumes that sampling grammar as extensively as possible is going to provide the best basis for wide-ranging prediction of real-world performance. (p. 3)

Harris and McCann (1994) conclude that this over-reliance on grammar tests results from the fact that it is easier to test tangible knowledge of grammar than to assess performance.

Harris and McCann (1994) also discuss the dichotomy between what happens in the teaching-learning process and evaluation when they argue that even though many teachers emphasize and prioritise communication in their classes, the tests that they use to evaluate their student progress and performance are mostly grammar-based. They point out that, "This has a very negative

washback effect on students. They quite naturally come to think that while speaking and listening are good fun, what really matters is grammar" (p. 32).

1.3 Testing purposes

The purpose of a test can be a key factor that determines every other aspect of how assessment is conducted and can lead, at the same time, to a good or a bad effect of the test on the students' learning process. On a general basis, Shepard (2002) distinguishes at least four different purposes for assessment in the educational field today.

At first, Shepard (2002) mentions *classroom assessment.* This type of assessment, either formal or informal, is an integral part of the teaching and learning process whereby not only student performance, weaknesses and progress can be evaluated but also teaching practices, usefulness of the material and effectiveness of the course in general.

Secondly, Shepard (2002) points out that *selection and placement tests* may be used to recognize students for gifted and talented programmes, to provide services for students with disabilities, or for college admissions. In the latter case, professional standards require that the results of such tests should be

considered together with high school grades and recommendations, since all students are prompt to have some degree of error associated with them.

Thirdly, Shepard (2002) discusses *large-scale assessment*. It can be used to monitor programmes and evaluate trends; its content must be comprehensive and inclusive of all the curricular goals of the many participating states or nations since there is not a single national or international curriculum.

Last, Shepard (2002) refers to the *high-stakes assessment of achievement*. It is different from large-scale monitoring assessment in that, apart from having different consequences, it must be closely aligned with content standards and curriculum and must be more limited in the variety of formats and tasks included. He concludes that this type of assessment should respect the most stringent technical standards because test inaccuracies in these tests could cause harm to individuals.

1.4 Different types of tests

The purpose of assessment can lead to another classification of tests. The following division provided by Harris and McCann (1994) can be said to be

related not only to the purpose for which the test is given but also to the moment when it needs to be administered.

- Placement tests

This type of test, also called entry test, can determine the level at which a student will most successfully learn. This kind of test is usually administered to students when the objective is to form courses with an intended homogeneous level (Harris and McCann, 1994).

- Diagnostic tests

This type of test serves to discover which contents the student does not know or has problems with and will consequently need to learn. This kind of test can be of great help for the teacher to design a course if it is administered at the beginning of the year. It can also help the teacher plan some remedial work when this type of test is used to detect problems during the year (Harris and McCann, 1994).

- Progress tests

This type of test, also called achievement test, can be administered at different times of the course to assess students' progress. Unlike the diagnostic test, which only detects failures, this type of test should be designed to discover student´s level of success; in other words, what students have learnt. The progress test can perform a very important formative function since not only can it provide information about students' progress but it can also give feedback

which can help the teacher and students solve certain learning problems (Harris and McCann, 1994).

- Summative tests

Summative tests are administered at the end of the course. They are usually used as a high stake test to decide on very important aspects such as promotion of a student to the following course or selection of students for higher level courses. Unlike progress tests, which give feedback that can improve the learning process, summative tests only provide a mark that classifies students. All the information that this type of test can provide about students' progress and weaknesses comes too late since this test is administered at the end of a course or school year and what students get is only a final mark (Harris and McCann, 1994).

- Proficiency tests

Proficiency tests are those that serve to show how successfully a student can use a foreign language. They are also required by some potential employers since these tests are popularly known to have a high component of reliability and validity, although some tests may not assess what the employees will be needing in their potential job (Harris and McCann).

1.5 Characteristics of a good test

Certain characteristics can provide a test with fairness, effectiveness, and practicality, thus ideally fulfilling the objectives for which it is being administered. These characteristics can affect different aspects of the test such as its design, its administration, its analysis or correction, and its results.

Validity, practicality, authenticity, interactiveness, impact and reliability can all, in different degrees, contribute to test usefulness, allowing the teacher to see if the test really serves the purpose for which it is intended. Bachman and Palmer (1996) refer to these characteristics as qualities of language tests that are included in a model of test usefulness and point out that:

> [...] test usefulness provides a kind of metric by which we can evaluate not only the tests that we develop and use, but also all aspects of test development and use. We thus regard a model of test usefulness as the essential basis for quality control throughout the entire test development process (p. 17).

1.5.1 Validity

A test is said to be valid when it really evaluates what it is meant to. *What* is the key word that every teacher should bear in mind when they are designing a test

and intend to make it valid. This *what* refers to the linguistic content or the elements of the language that are being assessed in a language test. However, validity in language tests does not only depend on the linguistic content of the test but also on the situation or technique used to evaluate this content (Lado, 1961).

In reference to linguistic content, a test which is meant to assess vocabulary, but requires that students complete fifty percent of it with grammatical structures, is not considered to be valid (Harris and McCann, 1994). As regards the situation or technique, the relation between the way in which the linguistic content is taught and the form in which this is assessed in the test will directly determine the validity of the test. An example of this would be a vocabulary test including a task with sentences in isolation to be completed with vocabulary. This test will not be valid if the vocabulary that is being assessed was taught and practised in context and in a communicative way. As Harris and McCann (1994) pointed out, this would have a very negative washback effect on students.

The absence of validity may be firstly detected by the test takers, who will notice that there seems to be no connection between the way they were taught some linguistic content and the way in which this linguistic content is being evaluated. This fact indicates that there should be a correspondence between what and how students are taught and on what and how they are evaluated (Bachman

and Palmer, 1996). This can also lead to the mentioning of one of Bachman and Palmer's (1996) tenets of their philosophy of language testing, "Relate language testing to language teaching and language use" (p. 13). One of the two fundamental principles on which their approach to language testing is based reads, "The need for a correspondence between language test performance and language use" (p. 9). Bachman and Palmer (1996) enlarge on this principle stating that for a particular test to serve its intended purposes, the performance of such a test must correlate in verifiable ways to non-test language use.

Bachman and Palmer (1996) provide the following example to demonstrate how performance on a language test should be related to language use in specific situations other than the language test itself:

> Suppose that you need to develop a test to determine if undergraduate students of Spanish as a foreign language at a North American University are ready, in terms of their language ability in Spanish, to spend a year studying entirely in Spanish at a university in Spain. You may already know that one use of Spanish will be in comprehending academic lectures, so you initially decide to include a task in your test (p. 10).

The validity of a test can be measured from different angles in different ways. Alderson, Clapham and Wall (1995) clustered the different ways of validating a test into *external* and *internal validity*. They argue that external validity comprises studies that compare students' test scores with measures of their

ability obtained from outside the test. This external form of validation includes *concurrent validity* and *predictive validity*.

1.5.1.1 Concurrent validity

Concurrent validity, as is implied by the term concurrent, is a form of validation in which the results of the target test[1] are compared to some other measures at roughly the same time that the target test is given. These other measures can correspond to the scores of another equivalent test, of the student's self-assessment of the performance of the ability that is being tested, or of the teacher's assessment of the language ability that is being evaluated in the target test (Alderson, Chapman and Wall, 1995).

1.5.1.2 Predictive validity

Predictive validity differs from *concurrent validity* in that the collection of the external measures is made some time after the target test is given. This type of validation is generally aimed at *proficiency tests* and *achievement tests*. The measure information to prove validity of the latter can be obtained from the teacher's comments on a student's suitability into a particular class or course or

---

[1] Test to be validated

from the results of another test, given right after the student has started the course and before any learning has taken place. The latter measurement is seldom done since validation of achievement tests is rarely conducted by means of another test or simply because the institution does not allow this kind of research (Alderson, Clapham and Wall, 1995).

Alderson, Clapham and Wall (1995) argue that the main purpose of proficiency tests such as IELTS or TOEFL would be to detect students who might not be ready, due to their weak level of English, to study in an English speaking setting. They state that predictive validation would involve giving the students one of the two tests mentioned above before they leave their home country. Once they have arrived in the host country, they would be given another test on their ability to use the English language in that academic setting. A high degree of predictive validity would be marked by a high correlation between the scores of both tests.

However, this type of validation can encounter different problems (Alderson, Clapham and Wall, 1995). Firstly, students may improve their level of the target language after they arrive in the English speaking country. This fact would lower the predictive validity coefficient. Secondly, as in the case of concurrent validation, institutions are seldom likely to give another test to validate a proficiency test that was given, in another country. Thirdly and occasionally, it is improbable that all the students that sit for the proficiency test in their home

country will travel to the English speaking country due to the fact that some of them may have obtained a failing mark or simply because they could not travel. In such a case, validation could be conducted with only a part of the population that originally sat for the test, that population being mainly the students who obtained a better mark. Thus, when this happens, the *truncated sample problem*[2] arises, lowering the predictive validity coefficient (Alderson, Clapham and Wall, 1995).

Alderson, Clapham and Wall (1995) classify internal validity into *face validity*, *response validity* and *content validity*.

1.5.1.3 Face validity

Alderson, Clapham and Wall (1995) suggest that, "Essentially face validity involves an intuitive judgment about the test's content by people whose judgment is not necessarily 'expert'" (p. 172). They argue that, although this judgment usually refers to the whole test, attention might also be paid to specific poor items. Alderson, Clapham and Wall (1995) also argue that although this type of validity is usually dismissed by testers, an emphasis on it has been made since the rise of Communicative Language Testing (CLT). They mention

---

[2] "You can only use a part of the original test population in the validation" (Alderson, Chapman and Wall, 1995, p. 181)

that the advocates of CLT emphasize the fact that a communicative language test should resemble something that someone does in real life with the language (Alderson, Clapham and Wall, 1995).

When face validity is the measure chosen to determine if a test is valid, a simple question can be asked: does this test look valid? As face validity does not require the judgment of experts, that question can be answered by any person who intends to see if the test is measuring what it appears to measure. Some ways to collect data on face validity suggested by Alderson, Chapman and Wall (1995) are interviews and questionnaire to students.

1.5.1.4 Response validity

Alderson, Clapham and Wall (1995) argue that the information that can be collected on how test takers reason or what they think when they respond to certain test items can serve to measure the validity of the test. They mention two ways of collecting data when dealing with *response validity*: introspectively and retrospectively. Data can retrospectively be gathered through an interview with the test taker after the test has been administered. Information can also be introspectively gathered if the test taker thinks aloud while doing the test. Nevertheless, Alderson, Clapham and Wall (1995) state that, these two ways may have drawbacks. They argue that the interviews to test takers, in the case

of validation, should contain very general questions and should only be given when answers are not specific or clear enough. The negative side of the interview can emerge when test takers do not remember why they have answered in a certain way. When information is gathered introspectively there should be a silent observer that will collect the data. This silent observer will be the one that informs on the reasoning of the test and not the test taker himself or herself, thus modifying the real test situation (Alderson, Clapham and Wall, 1995).

1.5.1.5 Content validity

The Chartered Institute of Educational Assessors defines *content validity* as a measure of how closely the content of an assessment matches the content of the test's specifications[3] that was designed for this purpose.

The process of content validation is usually conducted by a group of experts that usually meet in a systematic way to analyse the content of the test to be validated. That analysis consists of the comparison of the content of the test with a statement of what the content should be. That content statement could

---

[3] "A test specifications provide the official statement about what the test tests and how it tests it" (Alderson, Clapman and Wall, 1995, p. 9).

be the test's specifications or a formal teaching syllabus or curriculum (Alderson, Clapham and Wall, 1995).

The most important difference between content validation and face validation would be who the judgement on validity is passed by. In the content validation process, it is a group of experts that measure the validity of the test, whereas the person who judges whether a test has face validity does not necessarily have to be an expert on the content or topic that is being evaluated. Another difference would be the systematic process involved in content validation, not present in the face validity process. These two main differences have made some people disregard face validity and consider content validity a more serious form to measure the validity of a test (Alderson, Clapham and Wall, 1995).

1.5.1.6 Construct validity

Alderson, Clapham and Wall (1995) excluded *construct validity* from the internal and external validity classification partly due to the complexity of this process and partly because of the belief that construct validity is a superordinate form of validity to which internal and external validity contribute.

In order to define construct validity one should first define the term construct. "The term *construct* refers to a psychological construct, a theoretical conceptualization about an aspect of human behaviour that cannot be measured or observed directly. Examples of constructs are intelligence, achievement motivation, anxiety, achievement, attitude, dominance, and reading comprehension" (Ebel and Frisbie, as quoted in Alderson, Clapham and Wall, 1995, p. 183). After defining the term *construct*, Ebel and Frisbie refer to *construct validation* as "the process of gathering evidence to support the contention that a given test indeed measures the psychological construct the makers intend it to measure" (p. 183).

Alderson, Clapham and Wall (1995) refer to this process of validation as the *operationalisation* of a theory. As a way to enlarge this definition, some similarities and contrasts between construct validity and content validity could be mentioned.

Firstly, in the process of construct validation, like in the process of content validation, a group of experts meet. But in the process of content validation these experts meet to compare the items of the target test to the test specifications or the syllabus, whereas in the process of construct validity experts meet and in order to make judgments, they are given some definition of the underlying theory (Alderson, Clapham and Wall, 1995). Another difference between content and construct validation would be that in the latter, other ways

are used to assess to what extent a test, which measures a construct, is valid. Some theoretical evidence can be gathered and then compared to the construct being measured, or the test to be validated can be compared to another test, or the test scores can be correlated with other test scores. In relation to test scores, Bachman and Palmer (1996) state that:

> Construct validity pertains to the meaningfulness and appropriateness of the *interpretations* that we make on the basis of test scores. In order to justify a particular score interpretation, we need to provide evidence that the test score reflects the area(s) of language ability we want to measure, and very little else (p. 21).

1.5.2 Authenticity

A characteristic that can be closely related to validity is *authenticity*. Alderson, Clapham and Wall (1995) refer to the fact that, although the concept of authenticity is usually used to justify the validity of a test, the concept itself is seldom defined. They also add that when this term is meant to describe a test, it is usually treated as a synonym of validity.

However, Bachman and Palmer (1996) treat this term separately from validity. They point out that the higher the degree of correspondence between the test

task and the TLU task (target language use task), the more authentic a test will be.

1.5.3 Interactiveness

Interactiveness is another test characteristic, which Bachman and Palmer (1996) make reference to. They state that test interactiveness is determined by the interaction between the test taker or language user and the test task or the TLU task. They argue that the degree of test interaction rises as the degree of engagement of the test taker's areas of language knowledge, metacognitive strategies, topical knowledge and affective schemata in the test task, increases. As test quality involves the test takers' characteristics, it is also bound to bring a direct washback effect on the test taker.

Bachman and Palmer (1996) provide the following examples to illustrate how actual test tasks differ in terms of their authenticity and interactiveness:

The first example (A) is from a hypothetical institution abroad in which some of the typists do not understand English very well, but have nevertheless developed a high level of ability to perform certain typing tasks in English. These typists find it difficult to engage in any sort of reciprocal language use in English, or to produce written text in English on their own. Nevertheless they are excellent typists and produce high quality typescripts, even from handwritten documents, which is the only task

required for their job. A screening test for new typists in this situation might involve simply asking job applicants to type from a handwritten document. If the applicants know that their on-the-job use of English will be limited to exactly this kind of typing, they will probably perceive the typing test as highly relevant to the job. Clearly, however, the test meets very few of the criteria for interactiveness, since it does not necessarily require the test takers to process the handwritten document as language. That is, a typist might be able simply to copy the letters and words, without processing the document as a piece of discourse. This example illustrates a test task which would be evaluated as highly authentic but low in terms of interactiveness.

We can use the same testing situation for a second example (B). Suppose that these same applicants were capable of carrying on 'small talk'' conversations in English about food, the weather, clothing, and so forth, and suppose that we tested them by interviewing them in English. If the topics in the interview were of interest to them, the interview might actually involve the same types of interactions involved in non-test conversation. If the scores from this interview were used to select individuals whose sole use of English was to type from handwritten documents, how would this example rate with respect to authenticity and interactiveness? This task would probably be judged to be relatively low in authenticity due to the lack of relevance of the test task to the TLU[4] tasks. On the other hand, it would probably be rated relatively high in interactiveness, particularly if the interview format allowed the test taker a reasonable amount of control in selecting topics and influencing the structure of the interaction (pp. 27-28).

---

[4] Target language use

1.5.4 Reliability

A test is said to be reliable when it shows consistency in its results. This characteristic is present in the test when two tests, evaluating the same ability, are given to the same students on different occasions and the results obtained in both tests show to be the same.

However, reliability is far too difficult to be found as an absolute characteristic of the test due to different factors. One of the factors that can interfere in the consistency of test results can be time. A test taker is unlikely to do a test in the same way twice simply because, as Hughes (1991) stated, human beings are not likely to behave in exactly the same way on every occasion, even when the circumstances seem to be alike. Another factor that may intervene in the stability of results is learning. The test taker is bound to get different results in the second test if some learning took place between the first test and the second one. Still another factor that can affect the results of two identical tests that are given on different occasions can be memory. It can be quite difficult for a student to remember everything studied for a test, especially after the first test has already been administered. Lastly, a factor that can act as interference in the consistency of test results is scoring. A teacher may show some differences in the way the first and the second tests were marked.

The reliability of a test can be quantified in the form of a *reliability coefficient*, which will enable the comparison of the reliability of different tests. The ideal reliability coefficient is 1; a test with coefficient 1 is that test which would yield exactly the same results for a particular set of candidates regardless of when it happened to be administered. A test with a reliability coefficient of zero would give sets of results quite unconnected with each other, that is, the scores that the students actually obtained on a certain day would be no help in attempting to predict the score the students would get if they took the test the following day. Genuine test reliability coefficients are to be found between the two extremes of 1 and zero (Hughes, 1991).

Some methods can be used as a means to prove the reliability of a test. The different ways or methods used to prove reliability can be propped into *multiple administration methods* and *single administration methods*.

1.5.4.1 Test-retest method (multiple administration method)

In this method, the same test is given to the same subjects twice, on different occasions. The drawbacks that this method can present are manifold. Students may find the fact of having to do the same test twice boring or useless. If only a short time has passed between the first and the second administration, then

students are likely to recall the answers of the test. If, on the contrary, a long time has passed between the first and the second administration, some learning could have taken place or even some forgetting may intervene, thus leading to a difference between the results of the first and the second test (Hughes, 1991).

1.5.4.2 Alternate forms method (multiple administration method)

In this kind of method the same test, but with different forms, is administered twice to the same students. The same should be tested in both tests. One positive difference that this method has with regard to the test-retest method is the fact that, in this method, students are not likely to remember test items (Hughes, 1991).

1.5.4.3 Split half method (single administration method)

This method proposes only one test administration. The test is divided into two halves, which are supposed to be equivalent and the items have been ordered in terms of difficulty, a half may contain odd-numbered items and the other half even-numbered items. The purpose of having two halves is to get two sets of scores to obtain the reliability coefficient as if the whole test had been given twice. This method comes to be similar to the alternate forms method in that the

parts are equivalent (Hughes, 1991). Lado (1961) also referred to the split half method as the *chance-half method* and placed it over the test-retest method as more beneficial to students since with only one administration it can avoid affecting the student test performance.

1.5.4.4 Internal consistency method (single administration method)

Unlike the split half method, in this method there would be no division into halves but a single administration of the test with the application of a formula. Lado (1961), who came to call this method *inter-item consistency*, stated that it gives a measure of equivalence homogeneity, considering the population of students who passed and failed each item.

1.5.5 The relationship between reliability and validity

Alderson, Clapham and Wall (1995) state that the relationship between reliability and validity seems to be a simple one in theory but it is quite complex in practice. As reliability and validity are seldom found as absolute characteristics of a test, it is usually difficult for teachers to find a balance between these two attributes when designing a test. One of the two tends most often to be maximized at the expense of the other. The format of a test can

contribute to the degree of reliability or validity present in the test. Multiple-choice tests are likely to be highly reliable, whereas essays, for instance, tend to be more valid than reliable.

A good example used to illustrate the difference between validity and reliability, which involves a common bathroom scale, is proposed by Ratner (2010) in his article *The Predictive Model: Its Reliability and Validity*:

> If someone weighting 118 pounds, steps on the same scale, say, five consecutive times, yielding distant readings, say, 115, 125, 195, 140, and 136, then the scale is not reliable/precise. If the scale yields consistent readings, say, 130, then it is reliable, but not valid/accurate. If the scale readings are 118 for the five stepping on the scale, then the scale is both reliable and valid (p. 2).

1.6 Marking tests

"Marking is one of the most time-consuming parts of many teachers' job" (Harris and McCann, 1994, p. 55). Harris and McCann highlight the need to consider marking time when choosing test formats. Thus, the marking of a test can be closely connected to the test format, and this connection leads to another classification of tests: *objective* and *subjective tests*.

1.6.1 Objective tests

This type of tests can be beneficial for four reasons. First of all, all the students in large classes can do this type of test at the same time. The reduction in correction time can importantly benefit the teacher. Thirdly, these tests tend to carry a high degree of reliability. Lastly, the marking of these tests requires no experts, that is, they can be marked by anybody and the ones who will have to do this job can count on a unique marking key, which is not supposed to be subject to changes (Harris and McCann, 1994).

However, this type of tests can present certain drawbacks. The design of an objective test can be difficult and time consuming. It can often be noticed that the test designer may try to overcome these obstacles at the expense of validity. As a result, it can happen that the more reliable the teacher intends to make the test, the less valid it becomes (Harris and McCann, 1994).

1.6.2 Subjective tests

Subjective tests are those that require a specialist in the subject or topic to be corrected. An example of subjective tests is writing tasks such as essays or narrative compositions. They usually hold a high degree of validity and a low

grade of reliability. The degree of subjectivity can be reduced by using a marking scale for the correction of this type of tests, thus making them more objective. This marking scale will contain different points representing student performance (Harris and McCann, 1994).

1.7 Test results

Harris and McCann (1994) argue that on the one hand, test results will usually give a mark or score for each student but on the other hand, test results might simply tell the teacher that a student has passed or failed the test. Interpretation of test results can lead to two ways of making decisions about which students have passed or failed the test, *norm-referencing* and *criterion-referencing*.

1.7.1 Norm-referencing

In norm-referencing, the decision about failing or passing a student comes from a comparison of the results obtained by all the students who have done the same test. Students' marks are placed on a scale. The highest marks will go on one extreme and the lowest will go on the opposite extreme, leaving all the other marks in between the two extremes. Each mark will be represented by a percentage on the scale, according to the number of students that got that mark

in the test. The decision about which students have passed and which students have failed will depend on these percentages. This system shows to be quite arbitrary and unfair since the passing or failing will result from the quantity of students passing or failing the test and not from student performance in the test. This way is commonly used in public examinations and entrance exams (Harris and McCann, 1994).

1.7.2 Criterion-referencing

Contrary to norm-referencing, criterion-referencing leads to a decision based on test performance. The use of this system gets the examiner to produce certain criteria defining the desired performance the students should conform to. Thus, with this system, a pass or a fail will result from student's ability and not from a comparison of marks (Harris and McCann, 1994).

2 Vocabulary

2.1 Definition

The term *vocabulary* has always presented some degree of vagueness in meaning in the field of language teaching and learning. Hatch and Brown (1995)

state that "the term *vocabulary* refers to a list or set of words for a particular language or a list or set of words that individual speakers of a language might use" (p. 1). Similarly, Ur (1996) points out that "vocabulary can be defined, roughly, as the words we teach in the foreign language" (p. 60).

2.2 All about the word

As stated in all the definitions of vocabulary provided, the term *vocabulary* can be used to refer to a *word* or *words*. The dilemma arises when the word or words are to be dealt with as the equivalence of the vocabulary of a language. As stated by Read (2000), "the word is not an easy concept to define, either in theoretical terms or for various applied purposes" (p. 17). When he tries to define the term *word* he proposes certain distinctions for a better understanding of the term. The first distinction suggested by Read (2000) is the one between *tokens* and *types*, which refers to the count of words in a text. He states that the number of tokens in a text equals the total number of word forms, that is, the individual words that appear in the text more than once should be counted every time they appear. However, he suggests that the number of types equals the total number of the different word forms; therefore, a word that is repeated should be counted only once. In relation to tests tasks and this distinction, Read argues that:

When I set my students the task of writing a composition of 500 words, they and I understand that this refers to the number of tokens (or running words, if you like),

many of which – especially words like *the*, *a*, *to*, *and*, *in* and *that* – are multiple occurrences of the same type (Read, 2000, p. 18).

In the presence of vagueness in meaning, not only of the term *vocabulary*, but also of the term *word*, both terms have been more precisely replaced, in many instances, by the term *lexis*. Richards (2000) makes reference to the term *lexis* when, in an attempt to find an answer to the question 'what is a word?', he argues that:

> The first idea that probably springs to mind is *words*, a formulation that is admirably adequate for the layperson. But for anyone interested in exploring the subtlety and magic of lexis, the term *word* is too general to encapsulate the various forms vocabulary takes (p. 1).

It is clear therefore that Richards (2000) considers the term *word* to be too general and contrasts it to the term *lexis*, which could be assumed to replace the terms *words* or *vocabulary*. Similarly, Lewis (1993) presents the term *lexis* in this particular *way w*hen he introduces the fundamental principles in his book *The Lexical Approach* and he refers to the nature of *lexis* as the most important difference in naturally occurring language, and the contribution that it can potentially make to language pedagogy. He, once again, uses the term *lexis* when he states one of the key principles of the Lexical Approach, "Language consists of grammaticalised lexis, not lexicalised grammar" (p. vi).

2.2.1 Types of words

Another distinction related to words suggested by Read (2000) presents two main categories, *content* or *lexical words* and *function* or *grammatical words*. "Traditionally, grammatical words belonged to the domain of grammar teaching, while the teaching of vocabulary was more concerned with content words" (Thornbury, 2002, p. 4).

Content words, considered to be the major word classes, include *nouns*, *verbs*, *adjectives*, and *adverbs*. They are also referred to as *open class words* since new *content words* can be added to the language and the words borrowed from one language to another usually fit into this category. Moreover, these are words that have semantic qualities that can be described and can also be modified with a series of affixes (Hatch and Brown, 1995).

Hatch and Brown (1995) point out that, unlike content words, *function words* are a closed set because new forms for these functions are often infrequent and they argue that they are called function words because their meanings show how to determine relations between words in utterances. This category includes *pronouns*, *determiners*, *prepositions*, and *conjunctions*. Unlike content words, which play a more semantic role in the sentence, *function words* contribute to the grammatical structure of the sentence. It is precisely because of this

function that these words also come to be called *grammatical words*. (Thornbury, 2002).

2.2.2 Word families

If someone is asked how many words he or she knows from a given text, some time may pass before the person is able to provide the answer. The problem arises when the text includes words that present different forms and the person doubts whether to consider the word with the different forms as one word or several words. The type of words that can be affected by this change of forms are the content ones. All those words that have different forms but carry a similar meaning are said to belong to the same word family (Read, 2000).

An example that shows this relation among the words with different forms belonging to the same word family is *open*, *opens*, *opened*, *opening*, *openly*, *opener*. The first word in the example group is the unit to which other forms were added. Read (2000) calls this unit *base form*, whereas Hatch and Brown (1995) and Yule (2010) refer to this unit as a *lexical morpheme.* Yule defines a *morpheme* as "a minimal unit of meaning or grammatical function" (p. 67). Therefore, he makes a distinction between *lexical morphemes*, such as *woman*, *table*, and *happy* and *functional morphemes*, such as *in*, *where*, and *under*. Besides, Yule calls the *base form* a *minimal unit of meaning* or *free morphemes*

and the forms that can be added to these free morphemes *bound morphemes*. In addition, he divides the *bound morphemes* into *derivational morphemes*, also known as *suffixes* or *prefixes* as *–er* in *opener*, whose function would be to make new words or to make words of a different grammatical category from the minimal unit of meaning and *inflectional morphemes*, also known as *inflections*, like *–ed* in *opened,* whose function is to show grammatical aspects of the word.

2.2.3 Multi-word items

The dichotomy, already presented in the definition of vocabulary, arises from the need to know how to consider those meaningful units that include more than one word. In an attempt to clear out this problem, Ur (1996) states that "A useful convention is to cover all such cases by talking about vocabulary 'items' rather than 'words'" (p. 60). Additionally, Thornbury (2002) points out that these meaningful units are often called *lexical chunks* or technically known as *multi-word units*. Moon (1997) states that:

> A multi-word item is a vocabulary item which consists of a sequence of two or more words (a word being simply an orthographic unit). This sequence of words semantically and/or syntactically forms a meaningful and inseparable unit. Multi-word items are the result of lexical (and semantic) processes of fossilisation and word-formation, rather than the results of the operation of grammatical rules" (p.43).

Moon also states that, according to three important criteria, holistic multi-word items could be distinguished from other kinds of strings. These criteria are *institutionalisation*, which refers to the degree in which the multi-word item is recognized by the speech community as a meaningful unit that can reoccur as such in language; *fixedness*, which is the degree to which a multi-word item is a frozen sequence of items, that is, the items in the meaningful unit cannot occur in a different order; and *non-compositionality,* the degree to which the meaning of the multi-word item cannot be understood on a word-by-word basis.

Another analysis or classification of multi-word items can arise from the answer to the question formulated by Hatch and Brown (1995):

> How do we determine whether a phrase is really a lexical unit or simply a phrase made up of separate words. At one end of the continuum we might have free collocation and at the other end an unbreakable set (p.199).

Thus, a meaningful lexical unit or a multi-word item can be encountered in speech in different forms, such as *collocations*, *compounds*, *phrasal verbs*, *fixed phrases*, *proverbs*, and *idioms*. If all these forms of multi-word items were placed on a scale measuring the degree of institutionalisation, fixedness and non-compositionality, *collocations* would be at one end of the scale line holding the degree of more institutionalised, less fixed, and high compositionality; whereas *idioms* could be found at the other end of the scale line holding the degree of less institutionalised, more fixed, and high non-compositionality.

2.3 Vocabulary and learning

When Hatch and Brown (1995) refer to the distinction between *receptive vocabulary* and *productive vocabulary*, they state that "The most important point of the receptive/productive discussion is that it suggests that there are different ways to 'know' a word, that what is considered sufficient knowledge under one circumstance will probably not be sufficient under others" (p. 370). Haycraft (as quoted in Hatch and Brown, 1995) defines *receptive vocabulary* as, "words that the student recognizes and understands when they occur in a context, but which he cannot produce correctly" and *productive vocabulary* as, "words which the student understands, can pronounce correctly and use constructively in speaking and writing" (p. 370).

Nation (2001) points out that although reception and production can be seen on a continuum, there are other ways of viewing the distinction. The terms *passive* and *active* are sometimes used as synonyms of *receptive* and *productive vocabulary* because of the receptive skills, listening and reading and the productive skills, speaking and writing. However, this view has been widely objected. Belyayev (as cited in Hatch and Brown, 1995) criticized the dichotomy active/passive vocabulary, arguing that listening and reading should not be considered passive skills and therefore the vocabulary needed for those skills should not be regarded as passive, either. Nation states that:

Essentially, receptive vocabulary use involves perceiving the form of a word while listening or reading and retrieving its meaning. Productive vocabulary use involves wanting to express a meaning through speaking or writing and retrieving and producing the appropriate spoken or written word form (pp. 24-25).

2.4 Vocabulary and teaching

The teaching of vocabulary can be closely connected to learning. Nation (2001) suggests that teachers should ask four questions when thinking about how to organise the activities to teach vocabulary, "1. What is the learning goal of the activity? 2. What psychological conditions does the activity use to help reach the learning goal? 3. What are the observable signs that learning might occur? 4. What are the design features of the activity which set up the conditions for learning?" (p. 60).

The methodology used to teach vocabulary has shifted several times throughout history. Van Ek and Wilkins (as cited in Zimmerman, 1997) referred to the lack of attention to vocabulary in communicative language research or methodology. In relation to this, Zimmerman (1997) asserts that, "little explicit attention has been given to vocabulary in either theoretical or methodological publications about notional and functional syllabi" (p.13).

Different methodologies have been adapted and modified in accordance with the necessity to improve communication and to content criticism as well. The Grammar-Translation method, dating from the beginning of the nineteenth century, which proposed the study of bilingual word lists, was found inappropriate for focusing on the ability to analyse the language instead of focusing on the ability to use it. This discontent gave rise to the Direct Method toward the end of the nineteenth century. This method proposed interaction as the principal of natural language acquisition. The vocabulary that the learners were exposed to was supposed to be simple and familiar (Schmitt, 2000).

The criticism of so much oral exposure to the language with the Direct Method led to the Reading Method and to the Situational Method (Zimmerman, 1997). While using the Reading Method, the need to improve vocabulary skills to facilitate reading was highly emphasized by West in Great Britain. During the World War II the American military realized about their lack of people who were conversationally fluent in foreign languages. Thus, this need to quickly train soldiers in oral/aural skills gave way to the audio-lingual method (Schmitt, 2000). Zimmerman (as citied in Schmitt) argues that as the emphasis in the audio-lingual method was placed on teaching structural patterns, vocabulary was selected according to its simplicity and familiarity because of the need to make it easy.

The Audio-lingual Approach was followed by other meaning-based approaches, such as the Communicative Language Approach. Nevertheless, in the Communicative Language Approach, vocabulary has been given a secondary status; emphasis has been placed on issues of mastering functional language and how language connects together into larger discourse (Schmitt, 2000).

3 Assessing vocabulary

3.1 Historical overview of vocabulary testing

Nowadays, scholars in the field of language testing have a different perspective on vocabulary-test items of the conventional type, called discrete test items, which assess whether learners have knowledge of particular structural elements of the language. In the last thirty years of the twentieth century, language testers have progressively moved away from this type of items. The widespread rejection of the discrete test items has led to the adoption of the communicative approach to language testing. Today's language proficiency tests are not focused on assessing any particular knowledge but are based on tasks simulating communication activities that the learners are likely engaged in outside the classroom (Read, 2000). However, in reference to testing vocabulary in isolation or as part the language Nation (2001) claims that, "In some ways testing vocabulary is easier than testing grammatical knowledge or

control of discourse because the units to test are more obviously separate; it is not too difficult to identify what a word type is" (p. 344).

Read (2000) points out that not much attention is given to vocabulary testing and that this may be partly attributed to the overlapping of second language acquisition research and assessment. Thus, he ascribes some of the important work on how to measure vocabulary knowledge and ability to vocabulary acquisition researchers and not to language testers for two main reasons. Read believes that language testers work did not focus on vocabulary because they either took vocabulary tests for granted or, in the 1990s, were interested in more integrative and communicative measures of language proficiency.

"The history of vocabulary assessment in the twentieth century can be very much associated with the development of *objective tests*, especially in the United States" (Read, 2000, p. 75). Read argues that objective tests contain a series of test items, each with a single correct answer, that assess the different units into which the learning material was divided. They are known to be easy to score since no judgment by the scorer is needed to say whether an answer is correct or not. Objective tests find their origin in psychometrics, the science of mental measurement that influenced assessment, especially in the American schools after the First World War. Objective tests are also said to have progressively replaced traditional essay examinations from the 1930s till today. The importance of the role of vocabulary in objective tests gave rise to a great

amount of work in the 1920s and the 1930s to prepare lists of the most frequent words in English (Read, 2000). Standardized objective tests became the norm in the United States in the 1930s; this trend of testing vocabulary led to the creation of the *Test of English as a Foreign Language* (TOEFL). However, there has been a change on perceptions about how vocabulary should be tested since the rise of the communicative approach. Vocabulary is believed to be better measured in context than in isolation by many scholars. Thus, congruent with this thinking, in one of the most recent versions of the TOEFL, implemented in 1998, vocabulary items are embedded into computerized reading passages (Schmitt, 2000).

3.2 Vocabulary testing purposes

Vocabulary may be assessed for different purposes: *placement*, *achievement*, *diagnosis*, or *proficiency*. However, it is quite uncommon to find a language test that is used only to serve one of these purposes. At some time during the language course the teacher usually devotes part of a test to check if the vocabulary taught has been learned, that is, *achievement* of vocabulary; or to *diagnose* the problems the students may have with the vocabulary taught. Read (2000) points out that conventional vocabulary tests are most likely to be used by classroom teachers to assess progress in vocabulary learning and to diagnose problems in learning vocabulary. Read also points out that, researchers in language testing and those who undertake large testing projects

tend to be more concerned with the design of those tests that assess learners' achievement or proficiency on a broader scale.

Placement tests, which are used to determine the student level to be placed in a suitable language course, are seldom given in state run institutions (Harris and McCann, 1994). Placement vocabulary tests may be inside a *proficiency* test, which is of a commercial type and is in many cases taken to be admitted at university. The Test of English as a Foreign Language (TOEFL) can be considered to be one proficiency test that includes a wide range of vocabulary testing. In relation to this topic, Schmitt (2000) suggests that vocabulary tests that are part of proficiency tests should include the broadest range of words of all and should also include a range of words that can provide a fair evaluation of people from different countries and cultures, who may have different proficiency levels as well, since many universities usually rely on these proficiency tests to control their admissions.

3.3 Vocabulary and context

The role of context in the field of vocabulary can play a significant role mainly when, in assessment, one tries to answer the question whether vocabulary can be separated from other aspects of language proficiency. In the early years of objective testing, target words were usually presented in a decontectualised

way. However, the problem that testing vocabulary in isolation presents is that a word can have different meanings and can be used as more than one part of speech (Read, 2000). In reference to word meaning, Lewis (1993) argues that it is usually wrongly assumed that a word has a fixed meaning and to contradict this he presents the different meanings that can inhere in words. He draws a distinction among the *referential* meaning, that is the meaning that people know as *the meaning* or is also known as *core* meaning; *differential* meaning, the one produced by contrasting the language items; *connotational* meaning; *collocational* meaning; *pragmatic* meaning; *discourse* meaning; and *contextual* meaning.

When contextual meaning is concerned, the idea of text, written or oral, inevitably comes to mind. Ellis (as cited in Lewis, 1993) defines *context* in the following way:

> The 'context' of an utterance can mean two different things. (1) It can refer to the situation in which the utterance is produced; this is the 'situational context'. (2) It can refer to the linguistic environment – the surrounding language; this is the 'linguistic context'. Both types of context influence the choice of language forms and therefore have an effect on output (p. 80).

Lewis refers to the situational context simply as *context* and to the linguistic context as *co-text* or co-occurring language, which he considers more important than the situational context when the learning of language is concerned.

The topic of whether to present words in context or in isolation when teaching or assessing the language has produced several controversial arguments. Lewis (1993) claims that the two assumptions made by many teachers that words can only be learned when contextualised and that any form of putting the word in a sentence represents contextualisation can prove to be wrong since words may be accurately contextualised by learners in terms of their real world experience or imagination.

However, the construct validation studies by Corrigan and Upshur and Arnaud (as cited in Read, 2000) challenge the notion that vocabulary could be assessed as being something separate from the other components of the language. Evidence in their study also highlights the integral role that vocabulary plays in language ability, showing the strong relationship between vocabulary tests and the measures of reading comprehension. Read (2000) states that such findings give support to the view that vocabulary should always be assessed in context. On the other hand, Read points out that,

> As the research on the various members of the cloze family of tests shows, the mere we contextualise the assessment of vocabulary, the less clear it may be to what extent it is vocabulary knowledge that is influencing the test-takers' performance (p. 116).

3.4 Vocabulary measurements

Vocabulary tests can also be administered for measurement purposes. These measurement tests are usually used to measure two aspects of vocabulary that are *size* and *quality*. Schmitt (2000) argues that the purpose for testing vocabulary *size* or *how many* words a person knows is also referred to as *breadth of knowledge*, while the purpose of measuring *how well* target words are known is referred to as *depth* or *quality of knowledge*.

3.4.1 Measurement of vocabulary size

According to Read (2000), measurement of vocabulary size has been of a great importance in the educational field in different situations. One instance would be the reading programmes for people at different ages. He thinks that reliable estimates of words acquired by children at different age levels done by reading researchers could be useful information for making decisions about how many new words should be introduced in each unit of a learning programme. He also says that this kind of measurement can be useful to learn how much vocabulary a native speaker knows in order to compare it to the amount of vocabulary that a foreign language learner knows at the time of entering a school in a foreign country. An estimate of vocabulary size can also serve as a means to plan

educational programmes that will prepare students to meet their reading needs at university.

Schmitt (2000) states that, in order to give an estimate of a lower-level learner's total vocabulary size, frequency lists up to the 10,000-word level could be suitable used. Whereas the dictionary method would be more suitable for advanced learners and native speakers since it is necessary to sample from all the words in a language. In this *dictionary* or *sample rate method* certain words, for example the 5th word from every 10th page are systematically chosen from a large dictionary and fixed on a test. Then the percentage of correct answers is multiplied by the number of words in the dictionary to get an estimate of vocabulary size. However, this method appears to present several problems, such as the fact that the resulting total size estimate will highly depend on the size of the dictionary used or that the number of words finding their way onto the test compared to the total number of possible words is very low (Schmitt, 2000).

One of the most simple vocabulary tests to check whether a word is known or not is the *Yes/No* or *checklist* test. This test has become more popular since Anderson and Freebody (as cited in Nation, 2001) included some nonsense words in the test in order to measure the accuracy of the learners' responses; if a learner says he or she knows a nonsense word then they are overstating their vocabulary knowledge. An example of this is the *Eurocentres Vocabulary Size*

Test, which requires very little time to sit and is easy to interpret. However, as any other Yes/No tests, it does not give the chance to students to demonstrate knowledge of the meaning of the target words (Nation, 2001).

3.4.2 Measurement of vocabulary quality

The information obtained from the measurement of vocabulary size is not sufficient to determine whether a word is known or not known since there are different aspects that compose the knowledge of a word. In relation to this subject matter, Schmitt (2000) acknowledges that:

Lexical knowledge is complex, composed of a number of different components. In addition, vocabulary learning is incremental, and so the mastery of these aspects will tend to vary on a continuum stretching from "no knowledge" at one end to "complete knowledge and control" at the other. Thus, there is a great deal of latitude in what we might try to measure about the knowledge of a word, although in practical terms we will never be able to capture everything (p. 167).

In the discussion on which test best reflects how much knowledge the learner has of a word, Nation (2001) presented certain vocabulary items or tasks. Among these tasks he dealt with a multiple-choice task, a matching task, a true/false task, a definition completion task, a translation task. After discussing

how good each of those tasks can be to infer how much knowledge the learner has of the target word he concluded that, "to provide learners with the greatest chance of showing the vocabulary knowledge they have, it seems appropriate to use matching items with a sentence context" (p. 353).

3.4.3 Measurement Dimensions

3.4.3.1 Discrete vs. Embedded

A vocabulary test may have been designed to measure different aspects of the language. That aspect or mental attribute, or ability that is to be evaluated in a test is known as a construct. When a test measures the construct *vocabulary knowledge* distinctively from other components of the language is said to be *discrete*. An example of such kind of test would be a text with some words removed to be completed with the multiple-choice options. In this test only vocabulary knowledge is being tested. However, a test that included a text followed by some questions to answer about it would not be discrete but *embedded*. Even though it looks similar to the multiple-choice test, it is not measuring vocabulary knowledge exclusively. The construct that this test is primarily testing is reading ability. *Embedded* refers to, "A measurement of vocabulary which forms part of some other, larger construct" (Read, 2000, p. 9).

3.4.3.2 Selective vs. Comprehensive

Read (2000) states that the measure is said to be *selective* when the words to be evaluated in a vocabulary test were chosen in advance. These words may have been selected and then inserted in the text or the text may have been chosen first and then the words selected in that text. On the contrary, a *comprehensive* measure is the one that considers the whole vocabulary content of the input material or the test taker´s responses such is the case of a composition or a reading comprehension task with questions to answer. In reading comprehension tasks in which vocabulary is assessed comprehensively, readability formulas are used in the form of a calculation of the long words that appear in the text, since these words are known to be the less frequent and therefore, the most difficult ones (Read, 2000).

3.4.3.3 Context-independent vs. Context-dependent

As Read (2000) points out, contextualization has traditionally meant that a word is presented to test-takers in a sentence rather than in isolation. However, nowadays, the notion of context has been broadened to include whole texts and even discourse. Thus, this dimension, *context-independent – context-dependent*, goes beyond whether the words are presented in context or in isolation. Being *context-dependent* or *independent* has to do with:

To what extent the test-takers are being assessed on the basis of their ability to engage with the context provided in the text. In other words, do they have to make use of contextual information in order to give the appropriate response to the task, or can they just respond as if the words were in isolation? (Read, 2000, p. 11)

Read (2000) argues that the issue of context dependant also arises with cloze tests[5], in reference to which language testing researchers have debated whether they can mostly be answered correctly just by looking at the immediate context of the blank, or whether it is necessary to resort to information from the wider context of the passage. In relation to this, Read concludes that:

The degree of content dependence can be approached either as a characteristic of individual test items or as a property of the test as a whole. Generally speaking, vocabulary measures embedded in writing and speaking tasks are context dependent in that the learners are assessed on the appropriateness of their vocabulary use in relation to the task (p. 12).

3.5 Different forms of testing vocabulary

Thornbury (2002) makes the distinction between *receptive* and *productive tasks*. Receptive tasks can include all those *decision-making* tasks which engage the student in *identifying*, *selecting*, *matching*, *sorting*, *ranking*, or

[5] Test items, "in which words are systematically deleted from a text and the test-takers´ task is to write a suitable word in each blank space" (Read, 2000, p. 12).

sequencing vocabulary. On the contrary, *productive tasks* are those in which the learners are required to incorporate the newly studied words into some kind of speaking or writing activity and they can be classified into *completion* of sentences and texts or *creation* of sentences and text. Thornbury points out that receptive tasks can become productive tasks by having the learners talk about their judgement.

Read (2000) refers to vocabulary class tests as those test that should assess the learners' progress in vocabulary learning and at the same time give them an incentive to keep studying vocabulary on a regular basis. Read also argues that some practical considerations in the choice of test items should be made at the time of designing class tests.

3.5.1 Types of test items

3.5.1.1 Matching items

Read (2000) states that, "The basic matching task requires learners to make a connection between target words and their synonyms or definitions. As such, it is a recognition rather than a recall task, focusing on basic word meaning" (p. 171). Read points out that in a recognition task, the test-takers are presented with the target word and are asked to show their understanding of its meaning,

whereas in a recall task the test-takers are provided with some stimulus in order to elicit the target word from their memory. Read suggests that one or two extra definitions should be added, he is making reference to a match-the-words-to-their-definitions task, to avoid a situation where the test-taker can get definition correct by a process of elimination without really knowing what the word means.

3.5.1.2 Completion items

Read (2000) states that "completion, or blank-filling, items consist of a sentence from which the target word has been deleted and replaced by a blank" (p. 173). Read argues that the function of the sentence is to provide a context for the word and perhaps a cue of how to use it. He also says that completion items, unlike matching items, are recall tasks rather than recognition ones since the learners have to supply the target words from memory. Thornbury (2002) points out that completion items are more generally known as *gap-fills* and that a distinction can be made between *open* and *closed gap-fills.* In the open gap-fill, the learner has to fill the gaps by drawing on their mental lexicon, whereas in the closed gap-fill the learners are provided with a choice of words.

3.5.1.3 Sentence-writing items

Read (2000) considers the *sentence-writing item* to be the simplest vocabulary task and states that:

This task can allow the learners to demonstrate several aspects of their vocabulary ability:

- whether they understand the meaning of the target word;

- whether they know how the word functions grammatically within a sentence and what its correct form is;

- whether they know how the word collocates appropriately with other words; and

- more generally, whether they can use the word 'productively' in their writing (p. 175).

Thornbury (2002) refers to sentence-writing items as *creation* tasks in which learners are required to create the contexts for given words.

3.5.1.4 Multiple-choice vocabulary items

This particular type of vocabulary test format is considered to be objective since no expert is required for its correction. Wesche and Paribakht (as cited in Read, 2000) pointed out that even though the multiple-choice items are and will continue to be popular among test developers for being so much convenient to administer, they present several deficiencies such as, 1. they are difficult to construct; 2. the learner may know another meaning for the target word, but not the one being tested; 3. the right word may be chosen by a process of elimination; 4. Students' knowledge of distracters may be tested instead of the knowledge of the meaning of the target word.

3.5.2 Cloze tests

This type of vocabulary test can be designed in different ways. One kind of cloze is the *standard cloze*, which "consists of one or more reading passages from which words are deleted according to a fixed ratio (e.g. every seventh word)" (Read, 2000, p. 101). Another type, a modified version is the *selective deletion* or *rational cloze*, in which the words to be deleted are deliberately chosen by the test writer. Another type would be the *C-test*, which includes some short texts with the deletion of every second half of every second word.

Read points out that this version seems to be the less promising as a specific measure of vocabulary.

It is worth mentioning that the standard cloze and the rational cloze can present two different formats. The *multiple-choice* format, in which the test taker is provided with the text containing the blanks from the words that were deleted and four-option items. Another format in which the cloze can be presented is the *fill-in the blanks* one. In this type of cloze the test taker has to complete the blanks which correspond to the deleted target words.

Chapter III

The study

1 Context

The present study was conducted in three schools of English in the neighbourhood of Florida, Vicente Lopez, Buenos Aires. These schools of English were the only ones which had First Certificate in English (FCE) courses in the neighbourhood of Florida.

The University of Cambridge defines the *FCE examination* as, "an exam for people who need to prove they can use everyday written and spoken English at an upper-intermediate level for work or study purposes" (p.1). As stated by the University of Cambridge, the *First Certificate in English*, also known as *Cambridge English: First*, is set at Level B2 of the Common European Framework of Reference for Languages. It is also stated by the University of Cambridge that B2 is the level needed to start working in an English speaking environment and to study at an upper intermediate level. The University of Cambridge expanded on this level stating that:

For example, at B2 level typical users can be expected to:

- understand the main ideas of complex pieces of writing

- keep up a conversation on a fairly wide range of topics, expressing opinions and presenting arguments

- produce clear, detailed writing, expressing opinions and explaining the advantages and disadvantages of different points of view (p. 1).

It is worth mentioning that all the FCE courses in the three schools of English that took part of this investigation were supposed to have been designed according to Cambridge ESOL[6] Examinations Regulations published by the University of Cambridge Local Examinations Syndicate (UCLES).

2 Participants

Fourteen-to eighteen-year-old students, preparing the FCE examination in schools of English in Florida, participated in the study. A small portion of the totality of the participants, 12.2 %, included nineteen- to forty- year-old students.

The participants of the present study were considered to be FCE candidates at the moment this study was conducted. That is to say that the participants taking part in this investigation had already registered with an authorized Cambridge ESOL centre for the purpose of taking the FCE examination in December 2010.

---

[6] English for Speakers of Other Languages

3 Methods

3.1 Content analysis of test results

Content analysis is defined by Ary, Jacobs & Razavieh (1996) as "a research method applied to written or visual materials for the purpose of identifying specified characteristics of the material" (p. 485). Content analysis of three vocabulary test results was carried out in the hope of discovering if the hypotheses which gave way to the present research proved to be right.

3.1.1 Procedure

A carefully arranged timetable was agreed on with the head of each school of English. The undergoing study included all the First Certificate in English (FCE) students that were present the class in which the researcher gave the test. This segment of the research work took a month and a half; it was started in September and was finished in the middle of October. This time of the year was purposefully chosen for two reasons. Firstly, the FCE students had already done a great deal of practice by this time and secondly, students and teachers

were not so busy as they normally are in November when the time of the exam is approaching.

FCE courses, containing between 8 to 12 students on average, were divided into three groups**.** Each group was given three tests. As the identity of the participants was meant to be anonymous, they were asked to write their first name and surname initial in each test to facilitate further correction and scoring of the tests and a possible interview. The three groups completed the same tests, but at different times, in the same class. All the members of a group did the three tests in the same order. The students were asked to complete each test at a time. They were given the next test only after they had submitted the previous one.

Table 1 shows the order in which the groups did the tasks.

Order	Group A	Group B	Group C
1st	Task i	Task ii	Task iii
2nd	Task ii	Task iii	Task i
3rd	Task iii	Task i	Task ii

The students had been told by their course teacher that they were going to be given some practice by a person coming from outside the institution. Neither the content of the test nor the purpose of the investigation had been revealed to the students by the time they were asked to complete the tests. Students were asked to read the rubrics of each test carefully. The tests completed by each course were corrected within a very short period of time to give way to the next step in the study, the interview.

3.2 Interview

A semi-structured response interview was chosen as another instrument that, together with the tests, provided the necessary data to successfully conduct the present study. The research interview has been defined by Cannell and Kahn as: "a two-person conversation initiated by the interviewer for the specific purpose of obtaining research-relevant information, and focused by him on content specified by research objectives of systematic description, prediction or explanation" (as cited in Cohen and Manion, 1994:271).

The participants to be interviewed were chosen subjectively. Such interview was conducted with those students whose results in the three tests showed a certain degree of divergence. The purpose of this interview was to know the students´ opinion about each test in particular. Above all, the interview served

as a means to support the findings obtained in the administration of the three tests.

3.3 Triangulation

In the present study, two instruments, the tests and the interview, were applied in combination to cross check the data obtained. The administration of the three tests was followed by the interview, which mainly aimed at supporting the findings obtained in the tests.

Ary, Jacobs & Razavieh (1996) state that "*triangulation* - the use of multiple sources of data, multiple observers, and/or multiple methods - is another technique that is used to enhance the probability that hypotheses and interpretations are credible" (p. 480).

The administration of the three tests was followed by the interview, which mainly aimed at substantiating the results obtained in the tests. It was the aim of this study to produce quantitative as well as qualitative data since, as Ary, Jacobs and Razavieh point out, "Quantivative approaches in the human sciences rely on a hypothetico-deductive model of explanation. Inquiry begins with a theory of the phenomena to be investigated", whereas "Qualitative inquiry

seeks to understand human and social behavior from the 'insider´s' perspective-
that is, as it is lived by participants in a particular social setting (for example, a
culture, school, community, group, or institution)" (p.476).

4 Instruments

4.1 Analysis of tests

The FCE student participants were tested on the same ten words in the three
tests. Those ten key words had been taken from the *Handbook for Teachers for
Examinations from 2008* published by the University of Cambridge Local
Examinations Syndicate (UCLES). Each test added up to 10 points.

4.1.1 Multiple-choice cloze test

This is a FCE Use of English part one task. It was taken and adapted from the
Fast Track to FCE Coursebook (Stanton and Stephens, 2002). Hughes (1991)
states that, "In its original form, the cloze procedure involves deleting a number
of words in a passage, leaving blanks, and requiring the person taking the test
to attempt to replace the original words" (p.63). As Read (2000) points out, the
cloze task can present two formats: the *standard fixed ratio cloze*, in which

words are deleted according to a fixed ratio as for example, every seventh or ninth word one word is deleted and the *rational* or *selective-deletion cloze*, where the test-writer deliberately chooses the words to be deleted. Read also makes a distinction between the *fill-in-the-blank cloze item* and the *multiple-choice cloze item*, which, as Jonz (as cited in Read) points out, can be marked more objectively than the fill-in-the-blank cloze. It is worth mentioning that the task chosen for the present study is a *four-option multiple-choice cloze* with 10 blanks; five blanks less than in the original version for the sake of marking.

In this task students had to read a text with 10 gaps and choose from four options (A-D) to complete the text. This task focused mainly on vocabulary, but the answer to be chosen needed to fit both the meaning and the grammar of the sentence. The task included words that have similar meaning such as *say* and *tell*, collocations such as *do business*, set expressions such as *you are welcome*, linking words such as *and*, phrasal verbs such as *turn on*, and prepositional phrases such as *look for*.

4.1.2 Matching task

This test task consisted of two columns. The left column contained 10 words and the right column contained 15 words, five of which acted as distractors. It was intended, with this number of distractors, to reduce the possibility of

guessing because as Heaton (1988) states, a larger number of options or distractors would reduce even further the element of chance.

Students had to match the words from the left column to their corresponding synonym in the right column. The 10 key words, which also appear in the multiple-choice cloze task and the alternate choice task, were distributed in both columns.

4.1.3 Alternate choice task

This task consisted of 10 statements which contained the definition of a word. In this task students had to say whether the statements were correct or incorrect. There was a big guessing factor of 50% that ruled this task. When this task was designed, it was taken into consideration that, "Assuming that the focus of the test is on knowledge of the target words, the definitions should be easy for the learners to understand" (Read, 2000, p. 172).

4.2 Analysis of the interview protocol

Question # 1

¿Cuál de los 3 ejercicios te pareció más difícil? ¿Por qué?[7]

This question aimed at the students' acknowledging which of the three tasks presented greater difficulties. The sub-question *why?* led to the elicitation of the reasons why the student had found one exercise more difficult than the other two.

Question # 2

¿Cuál de los 3 ejercicios te pareció más fácil? ¿Por qué?[8]

In this question the students were asked to mention which of the three exercises they considered to be the easiest and to give the reasons that supported their choice.

The following three questions focused on obtaining certain information, introspectively, in order to discover the method that the student applied to do

---

[7] Which of the three tasks did you find more difficult? Why?

[8] Which of the three tasks did you find easier? Why?

each exercise. The details mentioned in the answers of the next three questions also helped understand the reasons given in questions # 1 and # 2.

Question # 3

¿En el ejercicio "multiple choice cloze", cómo llegaste a elegir la opción correcta?[9]

The purpose of this question was to find out the methodology used throughout the task in an attempt to get to the correct answer.

The following options were given in those cases in which the students did not know how to answer the question or as a means to enlarge on the answer.

A. ¿Qué te ayudó a elegir la opción correcta - el significado de las palabras en el contexto o la función sintáctica de las palabras cercanas al espacio a

---

[9] How did you get to choose the correct option in the multiple-choice task?

completar; es decir, te ayudó el ver que el espacio en blanco estaba precedido o seguido de una preposición, por ejemplo?[10]

This question aimed to discover if the students had got to the correct answer by understanding the context surrounding the blank and if the proximity of grammatical words had helped them choose the right option.

B. ¿Te focalizaste más en las cuatro opciones o en el texto?[11]

This question aimed to determine whether the student had focused his attention more on the text, that is, the words in context or on the 4 options, that is, the words in isolation.

Question # 4

¿En el ejercicio de unir, cómo llegaste a elegir la opción correcta? ¿Qué hiciste para llegar a entender el significado de cada palabra? ¿Pensaste en la función sintáctica de las palabras; es decir, te fijaste si eran verbos o sustantivos?[12]

---

[10] What helped you choose the correct option - the meaning of the words in the context or the syntactic function of the words surrounding the blank; that is, did a preposition before or after the blank help you decide on the correct answer?

[11] Did you focus you attention more on the four options or on the text?

This question encouraged the students to introspectively describe the process that they employed to do the exercise. It was mainly directed to discover whether the students had been able to find the correct answer just with the isolated word given or whether they had had to imagine the isolated word in a bigger structure such as a sentence as an example in which that word could be used. When the students were asked if they had thought of a context where they could place the given word, they were indirectly being inquired if they had been able to associate the isolated given word with some previous knowledge of the word existing in their mind. In this question students were also asked if the syntax of the given word had helped them make the choice.

Question # 5

¿En el ejercicio de elegir la opción correcta, creés que sabías el significado de las opciones dadas?¿Creés que pudiste haber adivinado el significado de alguna palabra?¿De ser así, qué te ayudó a adivinar el significado?[13]

---

[12] In the matching task, how did you get to choose the correct option? How did you try to understand the meaning of each word? Did you think of the syntactic function of the words; that is, if they were nouns or verbs?

[13] In the alternate choice task, do you think you knew the meaning of the ten points in the task? Do you think you guessed the meaning of any words? If so, what helped you guess the meaning?

This question aimed to discover if the students really knew the meaning of the options given or if they used guessing strategies to answer the question.

Question # 6

¿Con cuál de los 3 ejercicios estás más familiarizado?[14]

Question # 7

¿Con cuál de los 3 ejercicios estás menos familiarizado?[15]

In these two questions, the degree of familiarity was checked with the purpose of finding out about the student´s relation to each task independently of this investigation.

Question # 8

---

[14] Which of the three tasks are you more familiar with?

[15] Which of the three tasks are you less familiar with?

¿Cuál de los 3 ejercicios practicaste más este año?[16]

Question # 9

¿Cuál de los 3 ejercicios practicaste menos este año?[17]

Questions # 8 and # 9 were directed to elicit which task/s the students had devoted more practice to during their FCE course. As questions # 6 and # 7, they aimed to discover the student´s relation to each task independently of this investigation.

5 Results of the study

5.1 Tests

Three tests containing the same 10 key words were given to a total of 98 FCE students. For the sake of doing the three tests, the 98 FCE student participants

---

[16] Which of the three tasks have you practised more this year?

[17] Which of the three tasks have you practised less this year?

were divided into three groups in each course containing between 8 to 12 FCE students.

5.1.1 Analysis according to the scores obtained in the three tests

Figure 1

Figure 2

Figure 3

Matching: Percentage of students according to score

- Score 2: 9.18%
- Score 10: 7.14%
- Score 3: 10.20%
- Score 9: 12,.24%
- Score 7: 8.26%
- Score 8: 14.28%
- Score 5: 15.30%
- Score 4: 8.26%
- Score 6: 15.30%
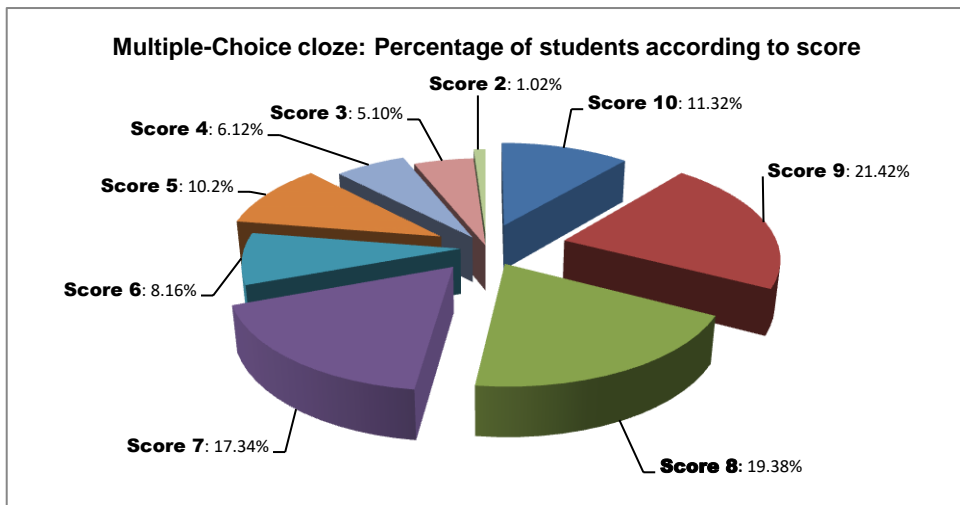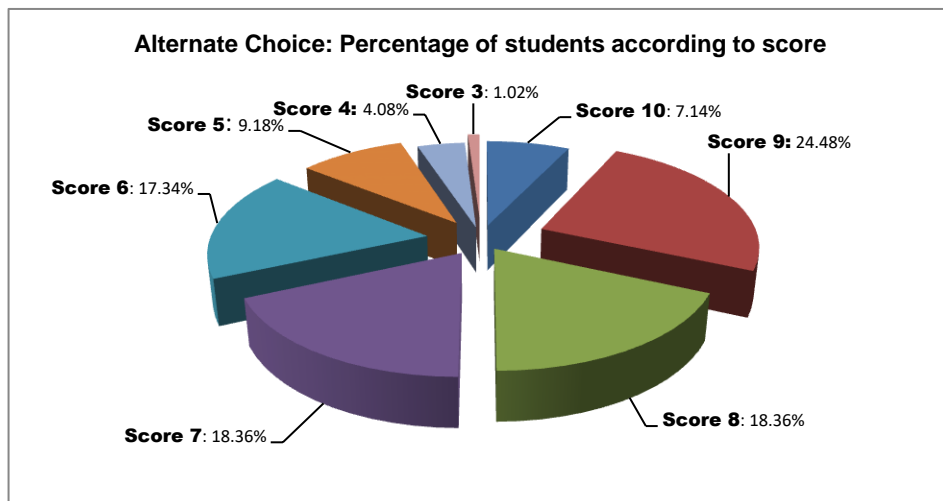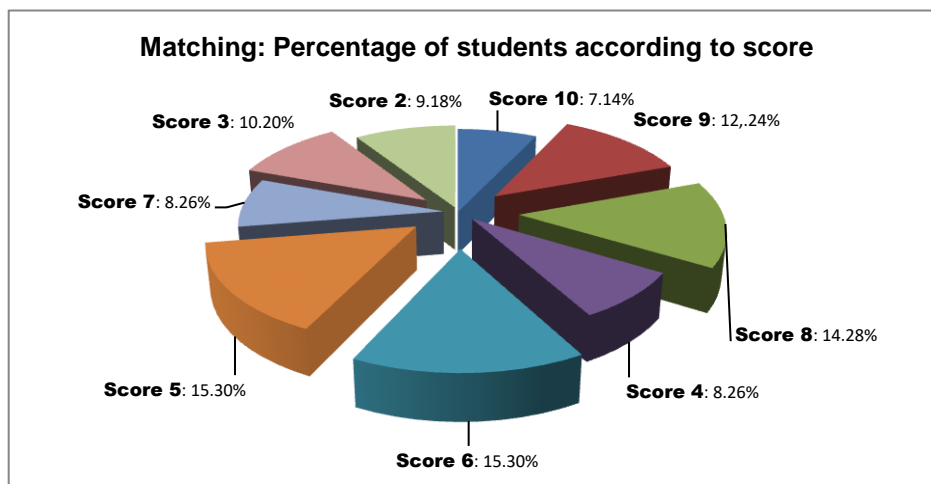
The results show that 11 students (11.2%) obtained the highest score (10/10) in the *multiple-choice cloze task*. Nevertheless, 7 students (7.1%) obtained 10 points in the *alternate choice task* and 7 students (7.1%) got the 10 points in the *matching task*. These findings would suggest that the *multiple-choice cloze task* may have been found easier than the *alternate choice task* and the *matching task* for presenting the biggest number of students with the highest scores. These results can be interpreted in terms of the degree of difficulty that the test-taker could have encountered in the test.

Figure 4

Students with the highest score

- 11.2% Multiple-choice task
- 7.1% Alternate choice task
- 7.1% Matching
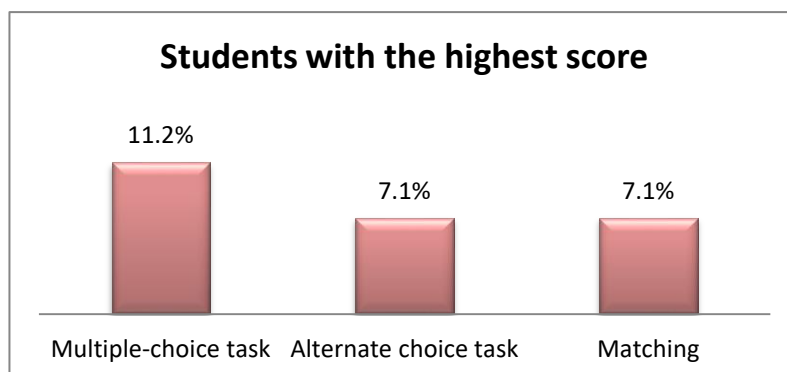
However, the percentage of students who obtained 9 points in the three tasks shows that the *matching task*, with 12 students (12.2%) obtaining 9 points, may not have seemed as easy to complete as the *multiple-choice task*, with 21 students (21.4%) who got 9, or the *alternate choice task*, with 24 students (24.5%) obtaining 9 points.

Moreover, score 8 and 7, which could also be considered within the range of high scores, was obtained by a bigger number of students in the *multiple-choice task* and in the *alternate choice task* than in the *matching task*. Results indicate that 19 students (19.4%) obtained 8 points in the *multiple-choice task* and 18 students (18.4%) obtained 8 points in the *alternate choice task*, whereas 14 students (14.3%) got 8 points in the *matching task*.

 Similarly in proportions comes to be the case of 7 points obtained in the three tasks; 17 students (17.3%) got 7 points in the *multiple-choice task* and 18 students (18.4%) obtained 7 points in the *alternate choice task*, but only 8 students (8.2%) obtained 7 points in the *matching task*.

Figure 5
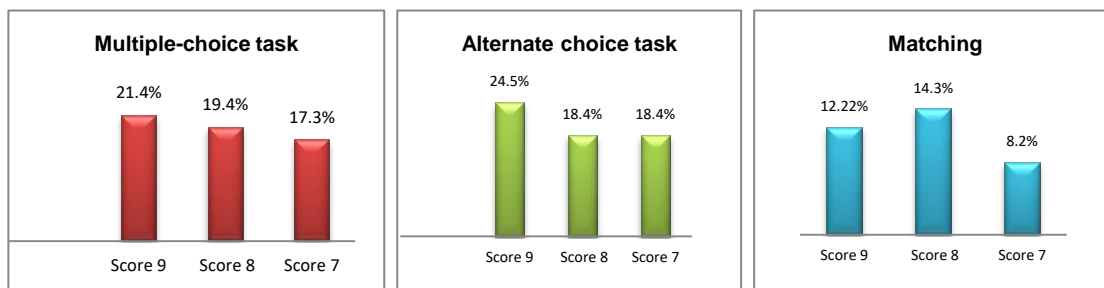
As table 2 clearly shows, the percentage of students who obtained the highest marks, from 10 to 8, in the *multiple-choice cloze task* (52%) resembles the percentage of students who obtained those marks in the *alternate choice task* (50%). These percentages can be contrasted to the percentage of students (33.67%) who obtained the highest marks, from 10 to 8, in the *matching task*.

Table 2

Multiple-choice cloze			Alternate choice		
Percentage	*# of students*	*Score*	*Percentage*	*# of students*	*Score*
	11	10		7	10
52%	21	9	50%	24	9
	19	8		18	8

Matching		
Percentage	*# of students*	*Score*
	7	10
33.67%	12	9
	14	8

Even more remarkable is the difference shown in table 3 between the percentage of students who got the range of scores that go from 9 to 7 in the *multiple-choice task* and in the *alternate choice task* and the percentage of students who obtained this range of scores in the *matching task*.

Table 3

Multiple-choice cloze			Alternate choice		
Percentage	*# of students*	*Score*	*Percentage*	*# of students*	*Score*
58%	21	9	61.2%	24	9
	19	8		18	8
	17	7		18	7

Matching		
Percentage	*# of students*	*Score*
34.7%	12	9
	14	8
	8	7

As Harris and McCann (1994) point out, once the students have taken a formal test and it has been marked, the test administrator will obtain results, which will give a score providing information about the student performance in the test. If the scores obtained by the participants are taken as a measure of the performance in the tests, from the comparison of the percentage of students who obtained high marks in the three tests, it can be inferred that a bigger

number of students seemed to have done better in the *multiple-choice task* and in the *alternate choice task* than in the *matching task*.

It is worth pointing out that a similar number of students got 6 points in the *alternate cloze task* and in the *matching task*, 17 students (17.3%) in the *alternate choice task* and 15 students (15.3%) in the *matching task* did 60% of the answers correctly, whereas only 8 students (8.2%) obtained 6 points in the *multiple-choice task*.

Interesting as it may seem, the number of students who obtained 5 points; that is, that did only 50% of the test correctly, is bigger in the *matching task* than in the *multiple-choice task* and the *alternate choice task*. Fifteen students (15.3%) got 5 points in the *matching task*, but a similar number of students, 10 (10.2%) in the *multiple-choice task* and 9 (9.2%) in the *alternate choice task* obtained 5 points.

Figure 6
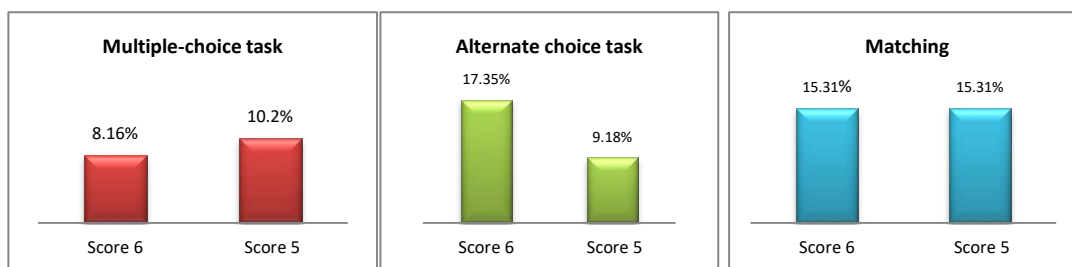
Once again, the number of students who obtained 4 points seems quite higher in the *matching task* than in the other two tasks, since 8 students (8.2%) got 4 points in the *matching task* but 6 students (6.1%) in the *multiple-choice task* and only 4 students (4.08%) in the *alternate choice task* obtained 4 points. Similarly, 10 students (10.2%) obtained 3 points in the *matching task*, whereas only 5 students (5.1%) obtained 3 points in the *multiple-choice task* and only 1 student (1.02%) got 3 points in the *alternate choice task*.

As can be noticeably seen in figure 7, 9 students (9.2%) obtained 2 points in the *matching task* but only 1 student (1.02%) obtained 2 points in the *multiple-choice task* and no students got 2 points in the *alternate choice task*.

Figure 7
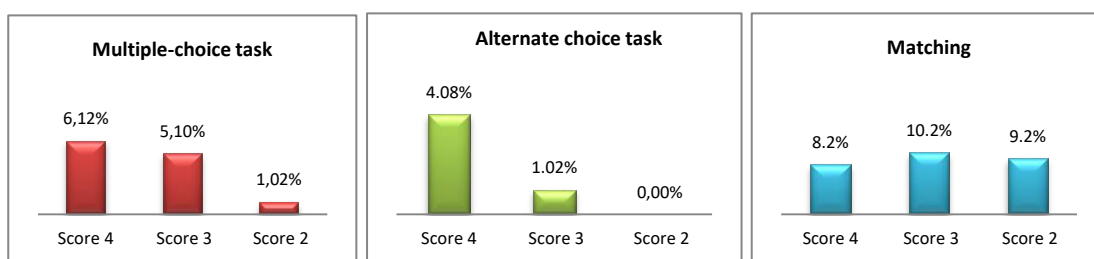
Once again, if the low scores are taken as a reference of the degree of difficulty that the test taker encountered in the test, it can then be concluded that the *matching task*, in which a bigger number of students obtained lower scores, resulted more difficult to do than the *multiple-choice task* and the *alternate choice task*, in which a bigger number of students obtained higher scores.

Table 4

HIGHEST SCORES (7 to 10)		LOWEST SCORES (1 TO 6)	
Task	# of students	Task	# of students
Multiple-choice cloze	68	Matching	57
Alternate choice	67	Alternate choice	31
Matching	41	Multiple-choice	30

As shown in table 5, a comparison among the three tests can be drawn in terms of number of students and scores obtained. As has already been pointed out, the same number of students, seven, obtained the highest score in the *alternate choice task* and in the matching task; whereas in the *multiple-choice task*, it was 11 the number of students who obtained the highest score. As from score 9, it can be noticed that the number of students tend to get smaller as the scores decrease in the *multiple-choice* and in the *alternate choice tasks*. This reduction in the number of students together with the decrease of scores appears to be absolute in the *alternate choice task* and almost absolute in the *multiple-choice task*, in which the number of students increases in score 5 and decreases again in score 4. However, it is worth mentioning that this reduction of the number of students with the decrease of the scores cannot be noticed in the matching task. This correlation between the decrease in the number of students and the decreasing scores in the *multiple-choice* and the *alternate choice tasks* can support what has already been pointed out about the fact that students did not seem to have encountered much difficulty in these two tests, whereas some

degree of difficult seems to have been present in the *matching task* since in this task there is not such a correlation as in the *multiple-choice* and the *alternate choice tasks*.

Table 5

Multiple-choice cloze

Number of students	Score
11	10
21	9
19	8
17	7
8	6
10	5
6	4
5	3
1	2

Alternate choice

Number of students	Score
7	10
24	9
18	8
18	7
17	6
9	5
4	4
1	3
----	2

Matching

Number of students	Score
7	10
12	9
14	8
8	7
15	6
15	5
8	4
10	3
9	2

5.1.2 Analysis according to order of task administration and score obtained

Results show that in the cases of the score increasing in the second task and the score increasing in the third task, the *alternate choice task* is the task that prevails over the other two tasks. Thus, 24 students obtained a higher score in the *alternate choice task* when this task was done in the second place, whereas 11 students obtained a higher score in the *multiple-choice cloze task* and 12 students obtained a higher score in the *matching* when these two tasks were done secondly. These results can mainly indicate that test performance tended to improve when the *alternate choice task* was done in the second place.

However, in the cases of decreasing score, it is worth pointing out that the *matching task* is the task that prevails over the other two tasks when it was done secondly. This is supported by the results that indicate that 19 students got a lower score when they did the matching in the second place, whereas 11 students obtained a lower score when they did the *multiple-choice task* secondly and only 4 students got a lower score when they did the *alternate choice task* in the second place. But in the case of the score decreasing in the third task, the same number of students, 17, obtained a lower score when they did the *matching task* and the *multiple-choice task* in the third place, whereas only 6 students obtained a lower score when they did the *alternate choice task* in the third place. According to the results that have just been mentioned, it

could be inferred that the *alternate choice task* was the task that appeared to be easier when done in the second and in the third place.

With regards to the highest scores obtained, it could be noticed that the difference of number of students who obtained the highest marks (7 to 10) between the *multiple-choice cloze* and the *alternate choice tasks* and the *matching task* is markedly bigger when the tasks were done in the first place, quite bigger when the tasks were done in the second place, but not so bigger when the task were done in the third place.

Table 6

Highest scores (7 to 10)

Test task	Multiple-choice cloze			Alternate choice			Matching		
Order	1	2	3	1	2	3	1	2	3
# of students	22	23	23	23	24	20	11	14	16

In relation to the students who obtained the lowest marks, it is worth pointing out that there is a big difference between the number of students who did the *matching task* in the first and second place and the number of students who did the *multiple-choice cloze* and the *alternate choice tasks* in that order. However, this difference is not that noticeable between the number of students who did

the *matching task* in the third place and the number of students who did the

multiple-choice cloze and the *alternate choice tasks* in the third place.

Table 7

<u>Lowest scores (1 to 6)</u>

Test task	Matching			Alternate choice			Multiple-choice cloze		
Order	1	2	3	1	2	3	1	2	3
# of students	21	21	15	10	8	13	11	8	11

5.1.2.1 The score trend present in the results

It was one of the purposes of the present study to discover if the successive

administrations affected the results obtained in the three tests in any way. Thus,

this investigation led to the encounter of the following results:

Interesting as it may seem, only 6.12% of the FCE student participants obtained

the same score in the three tasks. Five of the 6 students who got the same

score in the three tasks obtained 9 points, whereas the other student got 7

points. For the sake of this investigation, it is worth mentioning that out of the 5

students who obtained the same 9 points in the three tasks, only 1 of them did the same point incorrectly in the three tasks; 1 of them did the same point incorrectly in two of the three tasks; and the other 3 students answered different points incorrectly in the three tasks. It can also be interesting to mention that the only student who got 7 points in the three tasks answered incorrectly the same 2 points in the three tasks, whereas the other mistake he made did not coincide in any of the three tasks. To sum up, from the 8 mistakes made by the 6 students who obtained the same results in the three tasks, 46.9% of the incorrect points coincide in the three tasks, whereas 53.1 % of the incorrect points do not coincide.

Figure 8

6.12%
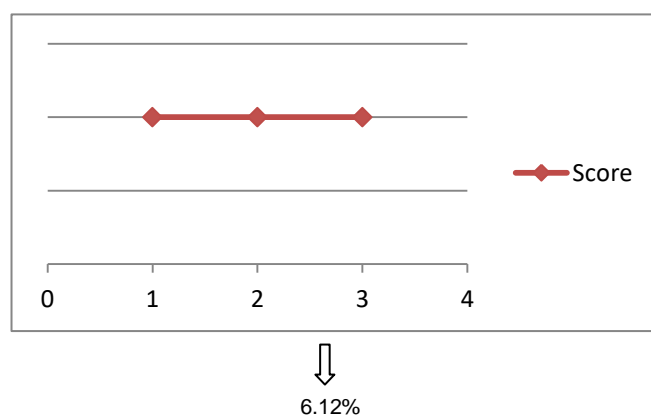
It is also worth pointing out that 10.2% of the FCE student participants obtained increasing results with the successive administrations of the tests. However, it was noticed that in 16.32% of the cases there was an increase of score that did not occur in a continuous form with the successive administrations of the tests.

In 7.14% of the cases result remains the same in the second test but increases in the third one and 9.18% of the cases show that result increases in the second administration but remains the same in the third. It is worth mentioning that the scores by 26.5% of the FCE students participating in the present study showed an increasing trend in the successive administrations of the three tasks.

Figure 9

10.2%

7.14%

9.18%
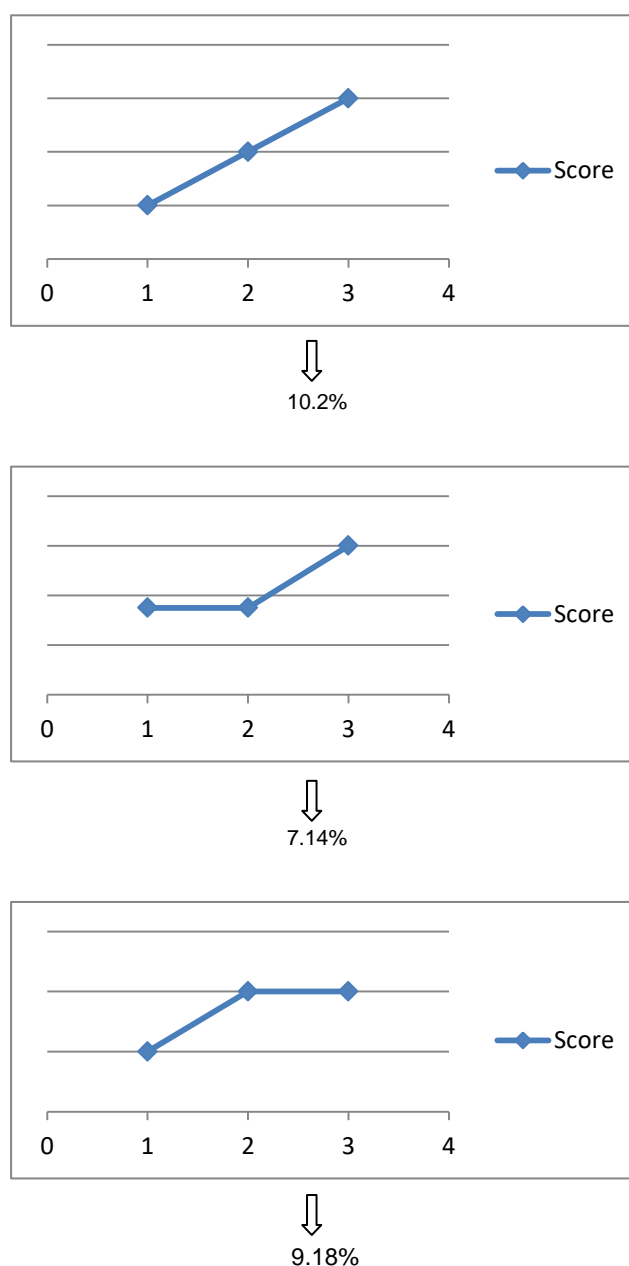
However, the score obtained by 28.6% of the FCE students increased in the second task but decreased in the third task. As shown in figure10, the scores obtained by 13.26% of the students increased in the second task but decreased in the third task; anyhow, the score obtained in the third task was higher than the one obtained in the first one. Moreover, the scores obtained by 3.06% of the students increased in the second task but decreased in the third and this score in the third task equaled the score obtained in the first task. Finally, the scores obtained by 12.24% of the students increased in the second task but it decreased in the third one, which was lower than the one obtained in the first task.

Figure 10

13.26%

3.06%

12.24%
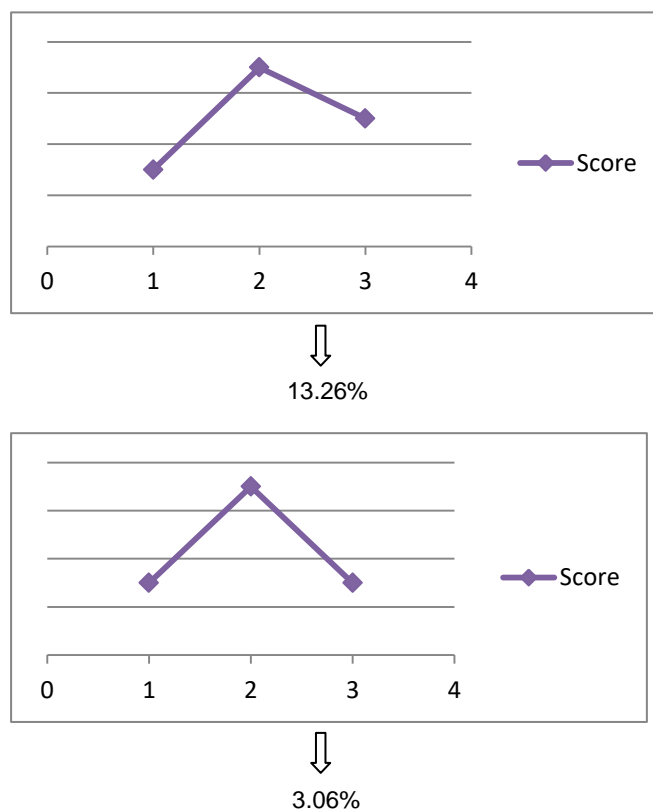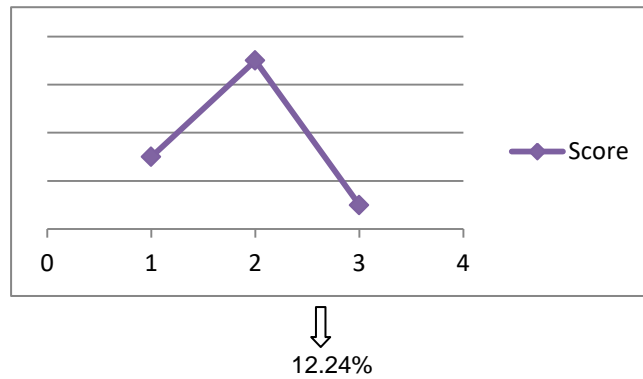
It is worth mentioning that 26.6% of the cases showed a decreasing trend of the score only in the second task administration.

As shown in figure 11 the score obtained by 9.18% of the students decreased in the second task but in the third task the score was higher than the one obtained in the first task. Furthermore, the score obtained by 5.10% of the students decreased in the second task but the score obtained in the third task was higher than the second and the same as the score obtained in the first task. Moreover, Figure 11 also shows that the score obtained by 3.06% of the students decreased in the second task but the score obtained in the third task remained the same as the second. Lastly, as can be seen in figure 11, the score obtained by 9.18% of the students decreased in the second task and then increased in the third task, but this third score was lower than the score obtained in the first task.

Figure 11
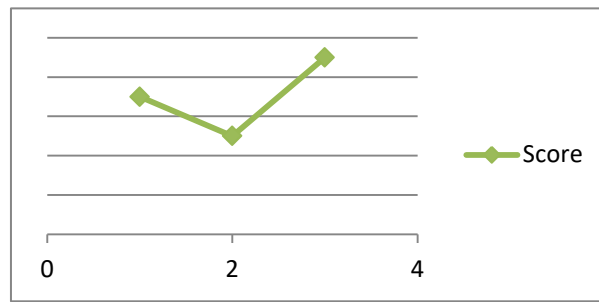
9.18%
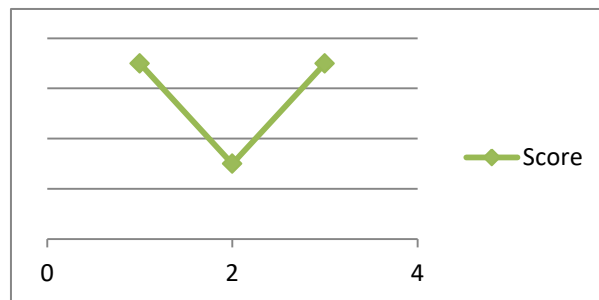
5.10%
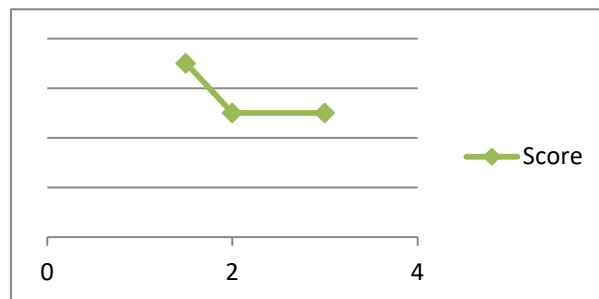
3.06%
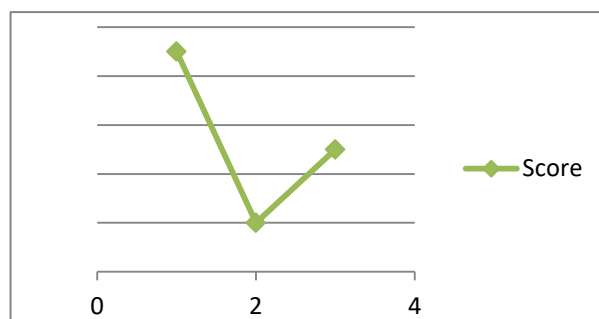
9.18%

As has already been pointed out, it is worth mentioning that 26.5% of the FCE students obtained a decreasing score only in the second task, but as figure 12 indicates, the score obtained by 8.16% of the students decreased successively with the second and third administration and the score obtained by 4.08% of the students remained the same in the second task but decreased in the third task.

Figure 12

8.16%

4.08%
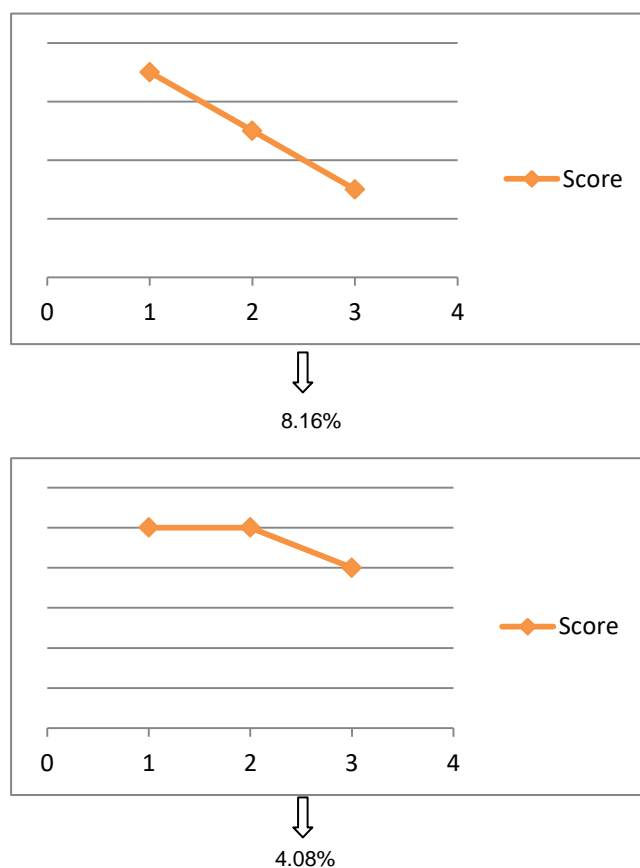
5.1.3 Analysis according to the answers in each task.

As surprising as it may seem, coincidences in the incorrect answers given in the three tasks seem to be really few considering the number of mistakes that have been made by the students in those three tasks. Only 1 question was coincidently answered incorrectly by 15 students in the three tasks, 18 students answered incorrectly the same 2 questions in the three tests, 7 students answered the same three questions incorrectly in the three tests, and only 4 students answered the same 4 questions incorrectly in the three test tasks.

Table 8

<div align="center">

<u>CORRESPONDENCE OF ANSWERS</u>

</div>

Multiple-choice cloze	*Matching*	*Alternate choice*
(1) A *recognized*	(2) N *to know – to recognize*	(I) *to recognize – to regard*
(2) C *results*	(7) A *to be as a consequence of – to result*	(B) *to result*
(3) D *as well as*	(9) H *together with – as well as*	(D) *as well as*
(4) A *leading*	(10) D *to cause-to lead to*	(H) *to lead to – to provoke*
(5) B	(1) K	(C)

conditions	circumstance- condition	a condition
(6) C likely	(5) J probable- likely	(G) Probably- likely to
(7) A provide	(3) G to supply – to provide	(E) to provide
(8) C predict	(8) C to forecast – to predict	(J) to prevent – to predict
(9) D cope	(6) M To manage – to cope with	(A) to cope – to manage
(10) B amount	(9) O quantity - amount	(F) an account – an amount

5.1.3.1 Answers to the multiple-choice cloze task

As the results of the present study indicate, the answer that most students did correctly in the *multiple-choice cloze task* is # 5, whose correct answer is B: *conditions*. Only 2 students chose option C: *cases* and no students chose A: *examples* or D: *instances*.

The second answer that was done correctly by most students is # 7, whose correct answer is A: *provide*. Option C: *prove* was chosen by only 2 students and D: *propose* was chosen by only 1 student.

The third answer that most students answered correctly is # 2, whose correct answer is C: *results.* Half of the students who did it wrongly; that is, 8 students chose option D: *happens*, 6 students chose B: *concludes* and only 2 students chose A: *recovers.* It is worth mentioning that in this question, the option with the most incorrect answers is the one that comes closest in meaning to the correct option. This can be supported by the definitions provided by the *Longman Dictionary of Contemporary English* (1995), *to result* means "to happen or exist as a result of something" and one of the definitions of *to happen* provided by this dictionary is "to be caused as the result of an event or action." Interestingly, both words share the idea of occurrence.

The question that was answered correctly in fourth place is # 10, whose correct answer is B: *amount.* Sixty percent of the students who did it wrongly chose option C: *average*, whereas 40% chose D: *account* and no students chose option A: *addition.* Like in question # 2, in this question, the option with the most incorrect answers is also the one that comes closest in meaning to the correct answer. According to the *Longman Dictionary of Contemporary English* (1995), one of the definitions of *average* is "the amount calculated by adding together several quantities, and then dividing this amount by the total number of quantities." Both words *amount* and *average* carry the idea of quantity in some way.

The fifth question that was answered correctly is # 3, whose correct answer is D: *as well as*. Thirteen of the 21 students who did it wrongly chose option C: *as long as*, whereas 7 students chose A: *as far as* and only one student chose B: *as soon as*. The choice of the incorrect answer C: *as long as* by the majority of the students who answered this question wrongly was not determined in this case, by a similarity in meaning to the correct answer as could be noticed in other questions. This can be justified by the fact that the options for this question are function words[18] and not content words as in other questions.

The question that comes in sixth place according to answers given correctly is # 4, whose correct answer is A: *leading*. Twenty-two students, out of 33 of the students who did this question wrongly, chose option B: *causing*, whereas 5 students chose D: *producing* and 3 students chose option C: *finishing*. Again, in this question, the option with the most incorrect answers is the one that comes closest in meaning to the correct answer. According to the *Longman Dictionary of Contemporary English* (1987), *to cause* means "to lead to or be the consequence of." It is worth pointing out that both words *leading* and *causing* share the same idea of resulting in something.

---

[18] "Unlike content words – nouns, 'full' verbs, adjectives and adverbs – they have little if any meaning in isolation and serve more to provide links within sentences, modify the meaning of content words and so on" (Read, 2000, p. 18).

Question # 1, whose correct answer is A: *recognized* was answered correctly by 65 students and incorrectly by 33 students. Almost half of the students who made a mistake in this question chose C: *registered* and the other half chose B: *regarded*. But only 2 students preferred D: *represented* as the right choice. As one might have expected, the wrong choices were made on options C: *registered* and B: *regarded* and not on option D: *represented*, chosen by only 2 students. Unlike the verb *to represent*, the verbs *to recognize*, *to register*, and *to regard* can be considered to contain stative senses when they refer to a single unbroken state of affairs (Greenbaum and Quirk, 1990).

Question # 6, with C: *likely* as the correct answer, was done correctly by 63 students. Most of the students who answered this question wrongly, 21 out of 35 students, chose B: *probably*, 11 students chose D: *possibly*, whereas only 3 students chose A: *surely* as the correct option. Once again most of the students who answered this question incorrectly chose the option that comes closest in meaning to the correct answer. The *Longman Dictionary of Contemporary English* (1995) defines *likely* as "probably", and *probably* as "an adverb used to say that something is likely to happen". From the definitions, it can be noticed that the decision choosing *probably* and not *likely* was determined by the close similarity in meaning that both words possess and the inability to recognize the syntactic difference between both words; *likely* should be followed by *to*, whereas *probably* should not.

Question # 8, whose correct answer is C: *predict*, was done correctly by only 45 students. Fifty three students, that is, more than half of the students who participated in this study did this question wrongly. Interestingly, most of these students, 52 out of 53 students, chose A: *prevent*. Only 1 student chose B: *pretend*, whereas no students chose D: *prefer*. According to the definition provided by the *Longman Dictionary of Contemporary English* (1995), that *to predict* means "to say that something will happen or that something will happen in a particular way" and that *to prevent* means "to stop something from happening," it can be inferred that *prevent* was chosen incorrectly by such a big number of students since both words, *prevent* and *predict*, share the idea of something happening.

Question # 9, whose correct answer is D: *cope,* was the answer that obtained the lowest points due to the fact that over 50% of the students, 57 out of 98, did it incorrectly. Twenty three of those students chose C: *manage*, 20 students chose A: *handle* and 14 students chose B: *face*. According to the *Longman Dictionary of Contemporary English* (1995), *to cope* means "to succeed in dealing with a difficult problem or situation," *to manage* can mean "to succeed in doing something difficult, especially after trying very hard," *to handle* means "to deal with a difficult situation or problem," and *to face* means that "if you face a difficult situation or it faces you, you must deal with it." From these definitions, it is not surprising to notice that these have been the questions with the biggest

number of incorrect answers; since the four option words carry the sense of difficulty in their meaning.

5.1.3.2 Answers to the matching task

Results also show that in the matching task the question that was answered correctly by the majority of the FCE students that participated in the present research is # 4, *quantity*, whose correct answer is O: *amount*. Noticeably, 15 students out of 19 students who answered incorrectly chose option E: *account*. If the definition provided by the *Longman Dictionary of Contemporary English* (1987) of *account* as "a sum of money kept in a bank" is to be considered, it could be said that this option was the most incorrectly chosen due to its proximity in meaning to the correct option. In fact, the three words *quantity*, *amount*, and *account* share the same idea of quantity in their meaning.

The second question that most students answered correctly is # 2, *to know*, whose correct answer is N: *to recognize*. Option B: *to face* was the option that most students chose incorrectly in this question. There must be something in the meaning of *to face* that led the students to this choice. One of the meanings of *to face* provided by the *Longman Dictionary of Contemporary English* (1995) is "to accept that a difficult situation or problem exists, even though you would prefer to ignore it." It might be assumed that this sense of acceptance and

ignorance implied in the meaning of the word *to face* could have influenced the wrong choice.

The next question with the highest percentage of correct answers is # 7, *to be as a consequence of*, whose correct answer is A: *to result*. Option D: *to lead to* is the one that presented the biggest number of incorrect choices. Like in other questions, from the definition provided by the *Longman Dictionary of Contemporary English* (1987) of *to lead to* as "to result in," it can then be deduced that option D was the option that most students chose incorrectly due to its proximity in meaning to the right answer.

Question # 5, *probable*, whose correct answer is J: *likely* was the one answered correctly in the fourth place. Surprisingly, option C: *to predict*, being a verb and not an adverb as *probable* or *likely*, was chosen by 60% of the students who answered this question incorrectly. However, it is worth pointing out that the idea of occurrence shared by the two words *probable* and *predict* might have led the students to the wrong choice. According to the *Longman Dictionary of Contemporary English* (1995), *probable* means "likely to exist, happen, or be true" and *to predict* means "to say that something will happen or that something will happen in a particular way."

In fifth place, question # 3, *to supply*, whose correct answer is G: *to provide* was correctly chosen by 63% of the students. And the option with the most incorrect choices is I: *to recover*. In order to understand why this option was the one that the biggest number of students chose incorrectly, it could be proper to consider the definitions provided by the *Longman Dictionary of Contemporary English* (1995) of *to supply* as "to provide people with something that they need or want, especially regularly over a long period of time," and *to recover* as "to get back something that was taken from you, lost, or almost destroyed." Thus, the idea of wanting to have something shared by both words may have induced the wrong choice of *to recover* as the synonym of *to supply*.

Question # 1, *circumstance*, whose correct answer is K: *condition* was chosen correctly by 61% of the students. Within this question, option F: *example* was the option with the highest number of incorrect answers. If the definitions provided by the *Longman Dictionary of Contemporary English* (1995) of *condition(s)* as "the situation in which people live or work" and *example* as "a thing, person, situation etc that you mention to show what you mean or to show that something is true" were considered, the choice of *example* instead of *condition* could probably be justified by the coincidence of both words meaning *a* situation.

Next, question # 8, *to forecast*, whose correct answer is C: *to predict*, was answered correctly by 51% of the FCE students. And the option with more

incorrect choices was L: *to prevent*. According to the *Longman Dictionary of Contemporary English* (1995), *to forecast* means "to make a statement saying what is likely to happen in the future, based on information that is available now" and *to prevent* means "to stop something from happening." Like the deduction already made between *to predict* and *to prevent* in the analysis of the answers to the *multiple-choice task*, it could be deduced that the wrong choice of *to prevent* instead of *to predict* was influenced by the idea of occurrence shared by the two words, which in this case is reinforced by the idea of the future time.

Questions # 6, *to manage*, whose correct answer is M: *to cope with* and question Nº 10, *to cause*, whose correct answer is D*: to lead to* were done correctly in fifth place. Coincidentally, in question # 6, the option with the biggest number of incorrect choices was B: *to face*, whose meaning comes very close to the word from the column on the right, *to manage*. One of the definitions provided by the *Longman Dictionary of Contemporary English* (1995) of *to manage* is "to succeed in doing something difficult, especially after trying very hard" and one definition of *to face* is "if you face a difficult situation or if it faces you, you must deal with it." It is worth mentioning that the idea of difficulty shared by both words may have led to the wrong choice. Similarly, in question # 10, the option with the biggest percentage of incorrect answers, A: *to result* is the one that comes closest in meaning to the correct answer, *to lead to.* According to the *Longman Dictionary of Contemporary English* (1995), *to lead to* means "to make something happen or exist as a result" and *to result* means

"to happen or exist as a result of something." The shared idea of occurrence between the two words might have had the students choose *to result* instead of *to lead to*.

All in all, it can be concluded that the distractors can, in some cases, be an absolutely negative influence on the choice of an answer in tasks like *multiple-choice cloze* or *matching*.

If the syntax of the answers is to be taken into account, table 8 clearly shows that the average of students that chose verbs incorrectly is proportionally bigger than the average of students that chose non-verbal phrases incorrectly.

Table 9

Verbs	Students that answered incorrectly		Non-verbal phrases	Students that answered incorrectly
To face	66		Account	29
To lead to	47		Amount	6
To cope with	38		As well as	22
To predict	26		Example	23

To provide	23	Condition	14
To prevent	32	Likely	6
To recover	26	Total	100
To recognize	5	X:	16
Total	280		
X:	31		

5.1.4 Comparison of the answers in the three test tasks.

Coincidently, question 8 C, *to forecast/to predict*, in the *matching* task and question J, *to prevent/to predict* in the *alternate choice* task, obtained the same number of correct answers and consequently, the same number of incorrect answers.

Likewise, questions 10, *amount*, in the *multiple-choice cloze* task with 78 correct answers and 4 O, *quantity/amount*, in the *matching* task with 79 correct answers and questions 9, *cope*, in the *multiple-choice cloze* task with 41 correct answers and 6 M, *to manage/to cope with*, in the *matching* task with 43 correct

answers were coincidently, very close to obtaining the same number of correct and incorrect answers.

Certain coincidences can also be noticed as regards the similar proportion of correct and incorrect answers given in the three tasks.

Both, questions 4 O, *quantity/amount*, in the *matching* task and F, *account/amount*, in the *alternate choice* task, obtained the biggest number of correct answers and consequently, the smallest number of incorrect answers.

That same pattern is present in questions 2, *results*, in the *multiple-choice cloze* task and 7 A, *to be as a consequence of/to result* in the *matching* task, since both questions obtained the biggest number of correct answers in third place.

Moreover, questions 1 K, *circumstance/condition*, in the *matching* task and C, *condition/example*, in the *alternate choice* task, obtained the biggest number of correct answers in fourth place.

As figure 13 clearly shows, there is quite a big difference in number between the questions that in the three tasks the students did correctly and the questions that the students did incorrectly. It can be noticed that only 2 of the 10 key

words, *amount* and *to* result, obtained a similar number of coincidences of correct answers in the three tasks. Remarkably, only 4 out of the 10 key words, *amount*, *to result*, *to provide* and *likely*, were, coincidentally, answered correctly in the three tasks by over half of the population that participated in the present study.

Figure 13

Coincidences of the answers in the three tasks

Amount; 63
To result; 61
To provide; 56
Likely; 52
To recognize; 44
Condition; 42
To lead to; 30
To cope; 22
To predict; 18
To lead to; 20
As well as; 18
To cope; 18
To predict; 16
Likely; 10
Amount; 3
To result; 2
To provide; 0
To recognize; 2
Condition; 0
As well as; 6

Correct   Incorrect
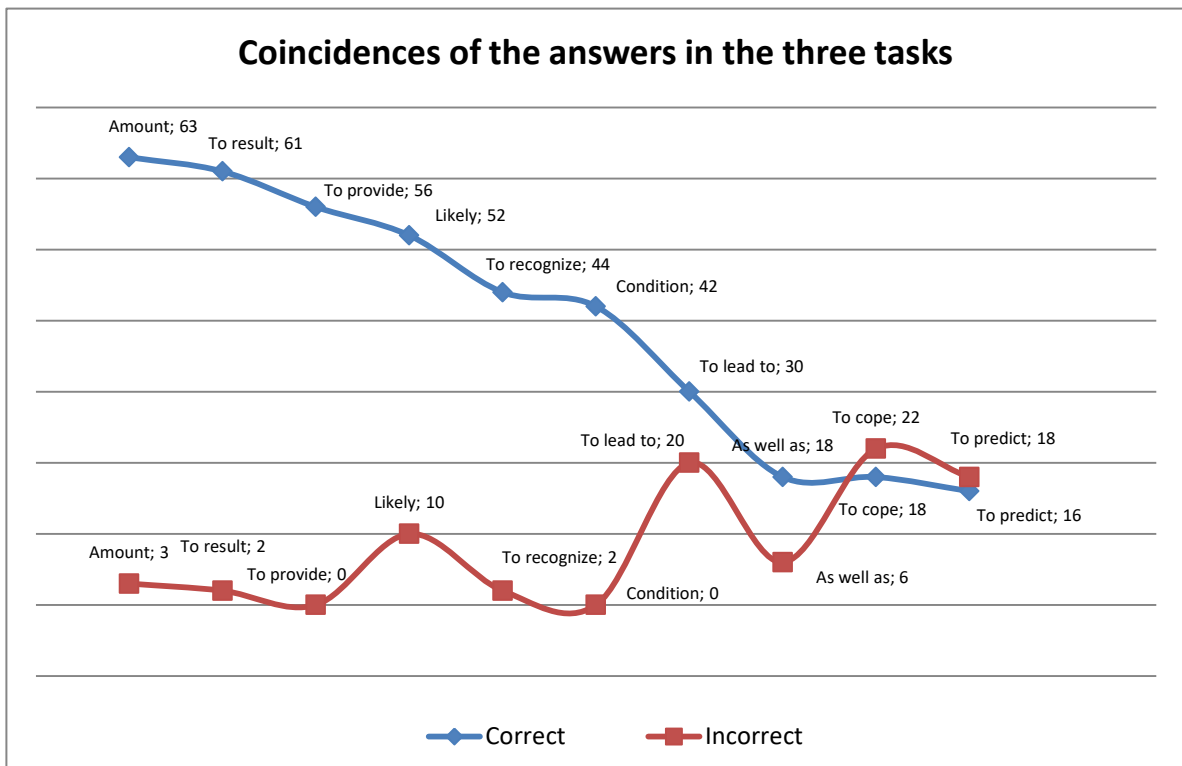
Ref:

1: *amount*, 2: *to result*, 3: *to provide*, 4: *likely*, 5: *to recognize*, 6: *condition*, 7: *to lead to*, 8: *as well as*, 9: *to cope*, 10: *to predict*

5.2 Interviews

The interviews were given to 33 of the 98 FCE students who participated in the present study. The interviewees were chosen for presenting a certain degree of divergence in the scores they obtained in the three tests. As the interviews aimed not only to discover the students´ opinions about the degree of difficulty found in each task but also the methodology they used to do the test, the results of the interviews will be presented in three sections.

5.2.1 Degree of difficulty found in the three tasks

The answers given to question #1 ("Which of the three tasks did you find more difficult? Why?") revealed the following:

As shown in figure 14, 24 out of the 33 students (72.72%) who were interviewed found the *matching task* more difficult than the other three tasks; 7 students (21.21%) found the *alternate choice task*; 1 student (3.03%) said that none of three tasks seemed to be difficult; and 1 student (3.03) found the *alternate choice task* as difficult as the *matching task*.

Figure 14

Task students found more difficult

1 student · 7 students · 1 student · 24 students

■ Matching ■ Alternate choice ■ Matching & Alternate choice ■ None of the three
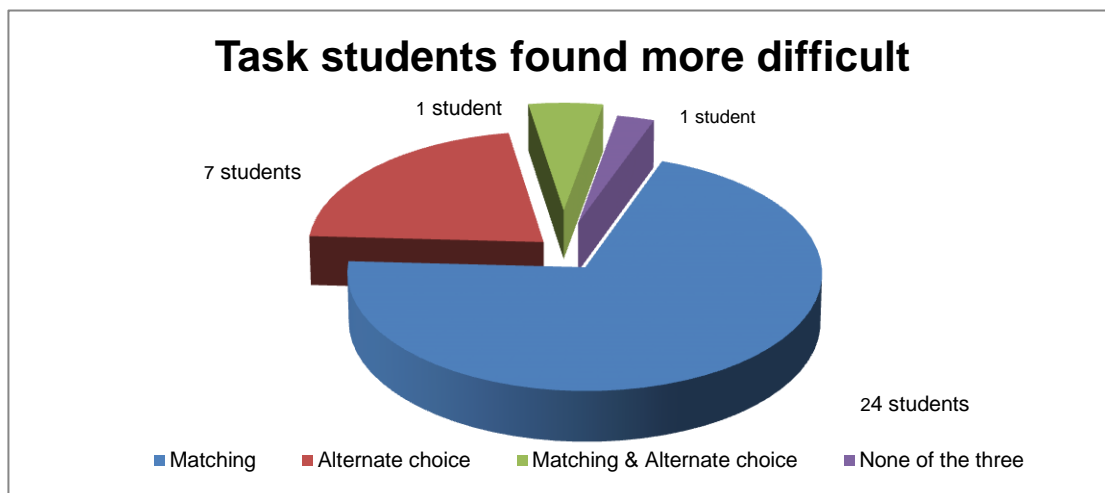
Out of the 24 students that pointed the *matching task* as the most difficult one, 14 students said that it was for having had to deal with the words in isolation without a context; 7 students said that the fact of having to match synonyms was confusing and that all the words looked similar; 2 students said that this task was found more difficult due to the order in which they did it, which was first; and 1 student said that the *matching task* looked more difficult than the others for finding more options in one of the two columns. Moreover, the students who found the *matching task* as difficult as the *alternate choice task* argued that they were confusing due to the lack of context in the tasks. Once again, the context seems to play a very important role at the moment of doing a task since, as has already been pointed out, the majority of the students that found the matching task more difficult related this difficulty to the lack of context in the task.

Out of the 7 students that found the *alternate choice task* more difficult, 3 students said that it was because of not presenting the words in a context as the *multiple-choice task* and the student who found this task as difficult as the matching agreed on this; evidently these 4 students were referring to the situational context for the reference they made to the *multiple-choice task*; 3 students said that they found this task more difficult because the definitions were confusing; and 1 student found it more difficult for having been the first task he did. It can be noticed the role of the context played a relevant role also in this task.

From answers given to question # 2 ("Which of the three test tasks did you find easier? Why?"), the following results were obtained:

As figure 15 shows, 24 out of the 33 students who were interviewed, 72.7%, found the *multiple-choice task* the easiest one; 8 students, 24.2%, thought that the *alternate choice task* was easier; and 1 student, 3.03%, found the *matching task* to be the easiest of the three test tasks.

Figure 15

Task students found easier

8 students   1 student

24 students

■ Cloze   ■ Alternate Choice   ■ Matching
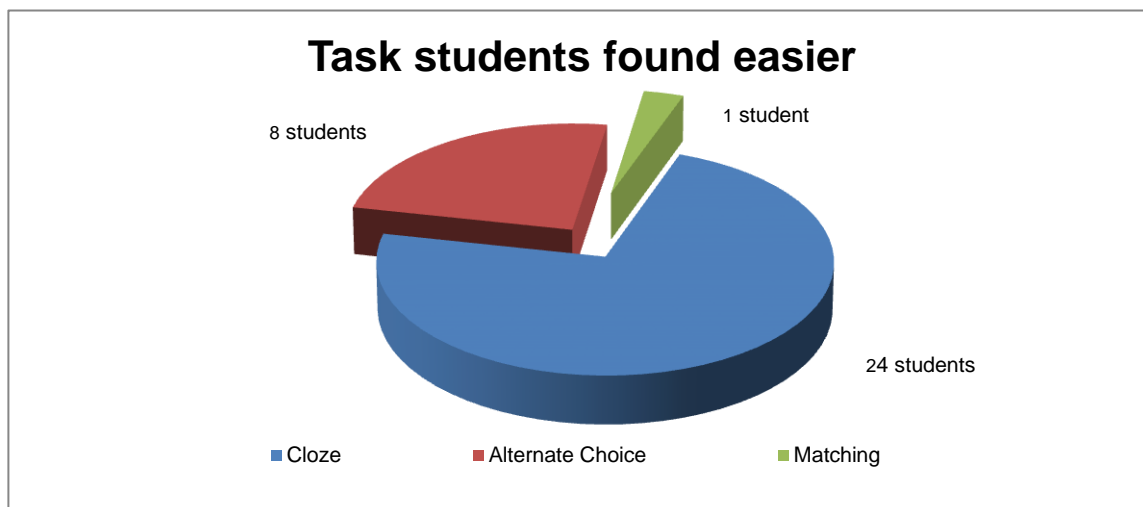
Out of the 24 students that found the *multiple-choice task* easier, 21 said they found it easier for having a context in which to place the right option, 5 of these 21 students said that they had also found it easier for having practised this task quite a lot during the FCE course and 3 students said that they had found this task easier because of the order in which they had done it; it is worth mentioning that these 3 students did the *multiple-choice task* in the third place. Interestingly, most of the students placed context as the main reason why they found the *multiple-choice task* easier than the other two tasks. Even though they did not specify in this question if they referred to the situational context[19] or the linguistic context[20], it can be assumed that it was mainly the situational context that made this task look easier than the others.

---

[19] It refers to the situation in which the utterance is produced (Ellis, as cited in Lewis, 1993).

[20] It refers to the linguistic surrounding or the surrounding language (Ellis, as cited in Lewis, 1993).

As regards the *alternate choice task*, out of the 8 students that found this task easier, 3 students said that it was because of having the definition of the word, 3 students said that they had found it easier for having a sentence or a context, 1 student said that it was because of the order in which he had done it, this student did the *alternate choice task* secondly, and 1 student said that the fact of having to say whether the statement was correct or incorrect made this test task easier.

Noticeably, only 1 student found the *matching task* easier than the other two test tasks. This student said that he had found it easier for the simple fact of having the synonyms there in the other column.

5.2.2 Techniques used to do the test tasks

The answers to question #3, which included sub-questions to let the student enlarge the way he used to do the task ("How did you get to choose the correct option in the *multiple-choice task*? A. What helped you choose the correct option, the meaning of the words in the context or the syntactic function of the

words surrounding the blank? B. Did you focus your attention more on the four options or on the text?") revealed the following:

In reference to question 3 A, the 33 students (100%) pointed out the fact that they could rely on a context helped them decide on the correct option and all of them also said that the syntactic function of the words surrounding the blank only helped them in some particular cases as when there was a preposition after the blank.

In relation to question 3 B, 14 students said that they had read the sentence up to the blank and immediately read the four options; 10 students said that they had read the whole text before reading the four options; 6 students admitted to having read the whole sentence before reading the four options; and 3 students said that they had read the whole paragraph before reading the four options.

The answers given to question #4 ("In the *matching task*, how did you get to choose the correct option, how did do try to understand the meaning of each word? Did you think of the syntactic function of the words; that is, if they were nouns or verbs?") revealed the following:

Out of the 33 students interviewed, 15 students said that they had to think of a context in which to include the word to be able to do the task; 7 students said that the used the guessing factor, that they did this test task at random; 5 students admitted that first, they matched the words known as a means of discarding and then tried to find the synonym of the unknown words; 3 students said that they had to repeat the words several times before matching them; 2 students said that they had to resort to translation into Spanish; and 1 student admitted that first, he looked at the syntactic function of the words. As interesting as it may seem, almost half of the students mentioned the context as a means to understand the meaning of a word.

In answering question # 5 ("In the *alternate choice task*, do you think you knew the meaning of the ten points in the test task? Do you think you guessed the meaning of any words? If so, what helped you guess the meaning?") the students said the following:

Out of the 33 students that were interviewed, 20 students said that they thought they knew the meaning of the word without having to guess; 6 students confessed that they had to think of an example in which to see the use of the word; and 7 students admitted to having used guessing strategies to do this task.

5.2.3 Familiarity with the three tasks

In question # 6 ("Which of the three tasks are you more familiar with?") the *multiple-choice task* resulted to be the test task that most of the students, 28 out of 33, said to be more familiar with; 3 students considered the *alternate choice task* as the most familiar to them; and 2 students said that they were familiar with both, the *multiple-choice task* and the *alternate choice task*. Interestingly, none of the students mentioned the *matching task*.

In answering question # 7 ("Which of the three tasks are you less familiar with?") 13 students said that they were less familiar with the *matching task*; 17 students said that the task they were less familiar with was the *alternate choice task*; 2 students said that they found both tasks, the *matching task* and the *alternate choice task* to be less familiar than the other one to them; and 1 students said that none of the three tasks resulted little familiar to him. Interestingly, the *multiple-choice task* was not mentioned to be the least familiar task to any of the students.

In questions # 8 ("Which of the three test tasks have you practiced more this year?") and question # 9 ("Which of the three test tasks have you practiced less this year?") 33 students (100%) said that the *multiple-choice task* was the task they had practised more during the year; 18 students said that they had

practiced the *alternate choice task* less than the other two tasks; and 15 students admitted that they had practised the *matching task* less than the other two during the year.

5.3 Triangulation of the results

As the purpose of this study was to discover to what extent test format can influence the results obtained in vocabulary tests, three different test tasks were chosen to be an instrument of the present research together with interviews to students, which provided information about the students' opinions and development of the tests. The triangulation of test results and the interviews will facilitate a comparison between the test outcomes and the account of the students' perceptions, in the hope of obtaining information that supports the results already analyzed in the present study.

As shown in table 10, 23 out of the 24 students who found the *matching task* the most difficult of the three obtained low scores, less than 7 points in this task; 6 students (25%) got only 2 points; 4 students (16.7%) obtained 3 points; 4 students (16.7%) obtained 4 points; 2 students (8.33%) obtained 5 points; 7 (29.16%) students obtained 6 points; and only 1 (4.2%) student obtained 8 points, which can be considered a high mark, in the *matching task*. As interesting as it may seem, the score obtained by 21 students in this test was

the lowest score that they obtained in the three tests. This finding may support these students' opinions about the degree of difficulty of this task in relation to the other two test tasks. As regards order, very similar figures to the *multiple choice task* were found; 11 out of the 24 students did the *matching task* first; 8 students did this task secondly; and 5 students did it thirdly.

As can also be seen in table 10, 4 of the students (57.14%) who found the *alternate choice task* more difficult than the other two obtained scores lower than 7; 1 student (14.28%) obtained 7 points; 1 student (14.28%) obtained 8 points; and 1 (14.28%) student got 9 points in this task; surprisingly this score was the highest that this student got in the three test tasks. As regards order, 4 out of the 7 students did this task first; 1 student did it secondly; and 2 students did this task thirdly. Once again, it can be pointed out that those 4 students' opinions about the degree of difficulty of the task may have been influenced by the order in which they had done the task.

Table 10

Tasks which were found more difficult

Matching		
24 students (72.7%)		
# of students	Score	Order
1	-2/10	2
2	--2/10	3
1	-3/10	1
2	--3/10	2
1	-3/10	3
1	-5/10	1
1	-5/10	3
2	--6/10	1
4	----6/10	2
1	6/10<	3
1	8/10<	1
3	---4/10	1
3	---2/10	1
1	-4/10	2

Alternate choice		
7 students (21.21%)		
# of students	Score	Order
1	5/10<	1
1	-5/10	2
1	7/10<	1
1	8/10<	1
1	+9/10	2
1	-4/10	1
2	6/10<	3

None of the three	Matching & Alternate
1 student (3.03%)	1 student (3.07%)

Ref:

+ The highest mark was obtained in the task.

-The lowest mark was obtained in the task.

<2nd score obtained in that task.

As table 11 shows, most of the students that found the *multiple-choice task* the easiest one (79.2%) obtained over 6 points in that task; 4 students (16.7%) got 6 points; and only 1 student (4.2%) obtained 4 points. It is also worth mentioning that the score obtained by 17 students (70.8%) was the highest score in the three tasks; the score obtained in this task by 6 students (6.12%) was the second in the three tasks; and the score obtained in this task by only one student (4.2%) was the lowest score in the three tasks. Another interesting point to mention can be the order in which the 24 students who found the *multiple-choice task* easier did the task; 11 students (45.8) did this task third; 6 students (25%) did it second; and 7 students (29.16%) did it firstly. Unlike the highest scores, which reflected students' performance, the order in which these 11 students did the task could have influenced their opinion about the degree of difficulty of the task.

Like in the *multiple-choice task*, most of the students who found the *alternate choice task* easier obtained high scores; from the 8 students that found this task easier, 6 students obtained over 6 points and 2 students 6 points. The score that 7 out of the 8 students obtained in this task was the highest in the three test tasks, whereas the score that 1 student obtained in this task was the second. As can be noticed in table 11, there is nothing as regard order that seems to be relevant.

Noticeably, the only student who found the *matching task* easier than the other two obtained the highest score in this task and did it in the third place.

Table 11

Tasks which were found easier

Multiple-choice cloze		
24 students (72.7%)		
# of students	Score	Order
1	+10/10	1
2	++10/10	2
3	+++10/10	3
2	++9/10	1
1	+9/10	2
2	++9/10	3
2	++8/10	1
3	++8/10<	3
1	7/10<	1
2	++7/10	2
1	6/10<	1
3	6/10-<<	3
1	4/10<	2

Alternate choice		
8 students (24.24%)		
# of students	Score	Order
1	+9/10	1
1	+8/10	1
1	+8/10	2
1	8/10<	3
1	+7/10	1
1	+7/10	3
1	+6/10	2
1	+6/10	3

Matching		
1 student (3.03%)		
# of students	Score	Order
1	+10/10	3

Ref:

+ The highest mark was obtained in the task.

-The lowest mark was obtained in the task.

<2nd score obtained in that task.

It is also worth mentioning that out of the 33 students who were interviewed only 11 students showed some coincidences of mistakes made in the three tests; 3 students (9.09%) did the same 3 questions incorrectly in the three test task; 7 students (21.2%) did the same 2 questions incorrectly in the three test tasks; and 1 student (3.03%) had the same mistake in the three tasks.

CHAPTER IV

Conclusions

As stated in the research question, the purpose of the present study was to find out to what extent test format influences test results in the assessment of vocabulary.

The information obtained from the analysis of the data collected through three test tasks and interviews to students seems to evince that the test formats used to assess FCE students´ knowledge of vocabulary influenced the results that those students obtained in the vocabulary tests.

As regards the first hypothesis, which stated that test performance is affected by the test format used for the assessment of vocabulary, it can be concluded that the scores obtained by FCE students in the three test tasks and their opinions about the test tasks proved it right. It was remarkable to observe that 92 out of 98 students (93.9%) obtained a different score either in two or the three test tasks. Moreover, a difference in scores obtained in the three tasks also showed that performance varied significantly in the three tests done by most of the students. The *multiple-choice task* seemed to be the task which showed the best performance since 11.32% of the students obtained the highest score (10 points in this task) whereas 7.14% obtained this score in the

alternate choice task and 7.14% got this score in the *matching task*. If the other highest scores (8 and 9) are taken into consideration, the *multiple-choice task* as well as the *alternate choice task* showed a better performance than the *matching task* since a big difference in the highest scores (8 and 9 points) can be noticed between the *multiple-choice* and the *alternate choice tasks* and the *matching task*. 21.42% of the students in the *multiple-choice task* and 24.48% of the students in the *alternate choice task* obtained 9 points, 19.38% in the *multiple-choice task* and 18.36% in the *alternate choice task* got 8 points, whereas in the *matching task* only 12.24% obtained 9 points and only 14.28% of the students obtained 8 points.

Moreover, the 33 students interviewed admitted that the three test tasks presented a different degree of difficulty; 24 students found the *multiple-choice task* to be the easiest and it is worth pointing out that 66.7% of these students had a very good performance in this task with 6 students obtaining the highest score in this test task. In contrast, 24 students pointed at the *matching task* as the most difficult task and it is worth mentioning that, coincidentally with their opinions, the majority of these students showed to have obtained low marks in this task, 17 students (70.8%) obtained less than 6 points and 6 students (25%) obtained 6 points in this task. This being much the case, it can be concluded that this first hypothesis proves right.

As far as the second hypothesis is concerned, which stated that test performance is not affected by the test format used for the assessment of vocabulary, it can be concluded from the findings in the present study, some of which have already been mentioned, that this assumption seems not to be right. The small number of students who obtained the same score in the three test tasks, 6 students (6.12%), and the students' opinion about the different degrees of difficulty presented in the tests tasks can greatly strengthen this conclusion. Moreover, the few coincidences in the answers given in the three test task, not only by these 6 students but on a general basis, indicate that student performance varied according to the type of task. As shown by the results already presented, only 3 students (3.06%) had the biggest number of coincidences in the three tests tasks; these three students answered incorrectly the same 4 questions in the three tasks. All in all, the second hypothesis can therefore be proved wrong.

In relation to the third hypothesis, which stated that test performance improves with the successive administration of the same or similar tests for the assessment of vocabulary, results seem to prove it wrong since a very low percentage of the student participants (10.2%) actually improved task performance with the successive administrations of the three test tasks. However, it is worth mentioning that even though in some cases (16.32%) the scores that students got did not increase successively, they either increased in the second test and remained the same in the third or remained the same in the

second and increased in the third. It is also worth pointing out that the scores of only 8.16% of the student participants decreased successively with the successive administrations of the three tests. To sum up, in the light of these findings, especially those which show such few students obtaining an increasing score with the successive administrations, the third hypothesis can prove to be wrong.

All in all, it can be concluded that test format can influence to a large extent test results in the assessment of vocabulary. These findings can be interpreted as evidence in support of Read´s claim that:

It is a well-established finding in testing research that the choice of test item to assess a particular skill or ability has an influence on the scores obtained. Therefore, we have to recognize two major sources of influence on test scores: the knowledge or ability represented by the construct and the testing task (2000: 96).

Limitations and suggestions for future research

Even though quite a big number of student participants took part in the present research, only a small portion of the population preparing the FCE examination was taken to be part of the investigation. The FCE students that participated in the research were all preparing this international examination in schools of English but, as can be assumed, there is a big number of FCE students preparing the examination at schools or with private teachers. In an attempt to

enlarge the scope of the population, students with different proficiency levels and of different ages could provide relevant information in future research on the assessment of vocabulary.

Given the importance of vocabulary in language learning as much as in language teaching, it could be suggested for future studies the inclusion of teachers' perception about this topic since the present investigation relied only on students' beliefs.

Another source of interest could be provided by a similar investigation into other areas, such as grammar. Through such study, the researcher could explore the field of assessment in the hope to find out if the order, in which the grammar tasks evaluating the same grammar points are given to students has any influence on the results obtained.

In the presence of the different tasks available to assess vocabulary knowledge, the only three tasks used in the present study could have limited the findings obtained in some way. The choice of the test tasks to be the instrument in the present study was made in an attempt to have a sample representing the different test tasks used to assess vocabulary. However, a greater choice of test tasks may widen the scope of findings in further research.

Works cited

Alderson, C., Clapham, C., & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.

Ary, D., Jacobs, L., & Razevieh, A. (1996). *Introduction to Research in Education*. Fort Worth: Harcourt Brace College Publishers.

Bachman, L. & Palmer, A. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.

Cohen, L. & Manion, L. (1994). *Research Methods in Education*. New York: Routledge.

Chatered Institute of Educational Assessors. (n.d.). *Assessment Validity*. Retrieved July 25, 2010, from http://www.ciea.org.uk/knowledge_centre/articles_speeches/general_articles/assessment_validity.aspx

Greenbaum, S. & Quirk, R. (1990). *A student's Grammar of the English Language*. Essex: Longman Group UK Limited.

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

Hatch, E. & Brown, Ch. (1995). *Vocabulary, Semantics and Language Education*. Cambridge: Cambridge University Press.

Heaton, J. B. (1998). *Writing English Language Tests*. Harlow, Essex: Longman.

Hughes, A. (1991). *Testing for Language Teachers*. Cambridge: Cambridge University Press.

Lado, R. (1961). *Language Testing*. New York: McGraw-Hill.

Lewis, M. (1993). *The Lexical Approach*. Hove: Language Teaching Publications.

Longman Group Limited (Ed.). (1987). *Longman Dictionary of Contemporary English* (2nd ed.). Essex: Longman.

Longman Group Limited (Ed.). (1987). *Longman Dictionary of Contemporary English*. Essex: Longman.

Moon, R. (1997). Vocabulary connections: multi-word items in English. **In N. Schmitt & M. McCarthy (Ed.),** *Description, Acquisition and Pedagogy* (pp. 40-63). Cambridge: Cambridge University Press.

Ratner, B. (2010). *The Predictive Model: Its Reliability and Validity*. Retrieved February 11, 2011 from http://www.geniqmodel.com/res/The-Predictive-Model-Its-Reliability-and-Validity.html

Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.

Shepard, L. (2002). *Standardized Tests and High-Stakes*. [Electronic version] **In J. W. Guthrie (Ed.)** Encyclopedia of Education (2nd ed. , Vol. 7). New York: Macmillan Reference USA.

Skehan, P. (1991). Progress in Language Testing: the 1990s. **In C. Alderson & B. North (Ed.),** *Language Testing in the 1990s* (pp. 3-21).

Thornbury, S. (2002). *How to Teach Vocabulary*. Essex: Pearson Education Limited.

University of Cambridge ESOL Examinations (2009). *Regulations 2010.* Retrieved September 10, 2010 from http://www.cambridgeesol.org/assets/pdf/resources/teacher/regulations.pdf

University of Cambridge Esol Examinations. *Cambridge English. First.* Retrieved August 7, 2011, from http://www.cambridgeesol.org/exams/ general-english/fce.html

Yule, G. (2010). *The Study of Language*. Cambridge: Cambridge University
Press.

Zimmerman, Ch. (1997). Historical trends in second language vocabulary
instruction. **In J. Coady & T. Huckin (Ed.),** *Second Language Vocabulary
Acquisition* (pp. 5-19). Cambridge: Cambridge University Press.

Appendix 1: The Tasks

Multiple-choice cloze

NAME: ……………………………… (Please write your first name and surname initial)

Read the text below and decide which answer A, B, C or D best fits each space.

GLOBAL WARMING

The Earth´s climate is a very complicated system. What´s more, it is now widely
(1) .... that human activity is having an effect on it. The pollution which (2)
.... from the use of oil and coal in industry, (3) .... the increased
use of private cars, is causing significant changes in temperature in many parts
of the world. These changes often have a knock-on effect on other aspects of
the climate, (4) .... to things like extreme weather (5) .... and
rising sea levels. Studying the changes which are taking place and predicting
those that are (6) .... to happen in the future is now a major area of
scientific research. The information which the scientists (7) .... is very
useful in helping governments to (8) .... the effects of climate change and
so be better prepared to (9) .... with them.

A much more serious problem, however, is how to prevent the situation from
getting worse. This depends on how quickly, and to what extent, the (10) ....

of pollution in the atmosphere increases. Although many countries have now agreed to try and limit the pollution they create, much more needs to be done. If no further action is taken, then temperatures are set to rise by about 0.2% per decade during the 21st century. Such a rate of warming is greater than anything that has occurred over the last ten thousand years.

1 A recognized   B regarded   C registered   D represented

2 A recovers   B concludes   C results   D happens

3 A as far as   B as soon as   C as long as   D as well as

4 A leading   B causing   C finishing   D producing

5 A examples   B conditions   C cases   D instances

6 A surely   B probably   C likely   D possibly

7 A provide   B progress   C prove   D propose

8 A prevent   B pretend   C predict   D prefer

9 A handle   B face   C manage   D cope

10 A addition   B amount   C average   D account

Matching

NAME: .... (Please write your first name and surname initial)

<u>Match the words in the left column with their corresponding synonym in the right column. There are five extra synonyms which you won´t need to use. Please write the correct letter next to the corresponding number.</u>

1- circumstance	A- to result
2- to know	B- to face
3- to supply	C- to predict
4- quantity	D- to lead to
5- probable	E- account
6- to manage	F- example
7- to be as a consequence of	G- to provide
8- to forecast	H- as well as
9- together with	I- to recover
10-to cause	J- likely
	K- condition
	L- to prevent
	M- to cope with
	N- to recognize
	O- amount

Alternate choice

NAME: ………………………….. (Please write your first name and surname initial)

<u>Say whether statements A to J are correct or incorrect. Write a (√) or a (x)</u>

A) *To cope with* and *to manage* mean to succeed in doing something difficult but if you managed to do something you succeeded in doing that something after trying very hard. ☐

B) *To result* means to happen or exist as a consequence of something that happened before. ☐

C) *A condition* is *an example* of a particular situation or problem.

D) If we say that you are doing something as well as something else, you are a specialist in both things. ☐

E) *To provide* something is to make sure that somebody gets what they need, especially by giving that something to them. ☐

F) *An account* of something is the same as *an amount* of something. ☐

G) If we say that something will probably happen we mean that something is likely to happen ☐

H) *To lead to* something means to *provoke* something. ☐

I) *To recognize* something is *to regard* something as *uncertain.* ☐

J) *To prevent* something from happening is *to predict* that something will happen. ☐

Appendix 2: The interview

1) ¿Cuál de los 3 ejercicios te pareció más difícil? ¿Por qué?

2) ¿Cuál de los 3 ejercicios te pareció más fácil? ¿Por qué?

3) En el ejercicio "*multiple-choice cloze*" ¿cómo llegaste a elegir la palabra correcta? ¿Qué te ayudó a elegir la palabra correcta?

   (En el caso que el alumno no conteste esta pregunta claramente lo ayudaré con las siguientes preguntas:)

   - ¿Te ayudó el significado de las demás palabras dentro del contexto?
   - ¿Te ayudó la sintaxis de las demás palabras próximas a la palabra clave (por ser preposiciones o artículos….)?
   - ¿Enfocaste tu atención sólo en las 4 opciones?

4) En el ejercicio "*matching*" ¿Cómo llegaste a elegir la opción correcta?

   (En el caso que el alumno no conteste esta pregunta claramente lo ayudaré con las siguientes preguntas:)

   - ¿Recurriste a entender el significado de cada palabra?

- ¿Recurriste a ver qué tipo de palabra era: verbo, adjetivo, sustantivo? ¿Esto lo hiciste al principio, durante el ejercicio o al final cuando chequeaste el ejercicio?

5) En el ejercicio "*alternate choice*" o "*correcto or incorrect*" creés que sabías el significado de todas las palabras de los 10 puntos? ¿Creés que adivinaste el significado de alguna palabra? ¿Qué te ayudó a adivinar el significado?

6) ¿Con cuál de los 3 ejercicios estás más familiarizado?

7) ¿Con cuál de los 3 ejercicios estás menos familiarizado?

8) ¿Cuál de los 3 ejercicios practicaste más este año?

9) ¿Cuál de los 3 ejercicios practicaste menos este año?

Appendix 3: Score obtained according to order of task administration

Score	Order of task	Matching			Multiple-choice cloze			Alternate choice		
		1º	2º	3º	1º	2º	3º	1º	2º	3º
10		-----	3	4	2	4	5	-----	4	3
9		4	4	4	9	7	5	12	6	6
8		3	5	6	4	5	10	5	7	6
7		4	2		7	7	3	6	7	5
6	Number of students	8	5	2	2	1	5	3	7	7
5		4	5	6	4	3	3	4	1	4
4		3	3	2	----	4	2	2	-----	2
3		3	4	3	4	----	1	1	------	----
2		3	4	2	1	----	-----	-----	------	----
1		----	-----	----	----	-----	-----	-----	------	----

Appendix 4: Answers to each task

MULTIPLE-CHOICE CLOZE																			
(1)				(2)				(3)				(4)				(5)			
C		*I*		*C*		*I*		*C*		*I*		*C*		*I*		*C*		*I*	
Answer	*Nº of students*	*Answer*	*Nº of students*	*Answer*	*Nº of students*	*Answer*	*Nº of students*	*Answer*	*Nº of students*	*Answer*	*Nº of students*	*Answer*	*Nº of students*	*Answer*	*Nº of students*	*Answer*	*Nº of students*	*Answer*	*Nº of students*
		B:	15			A:	2			A:	7			B:	22			A:	0
		C:	16			B:	6			B	1			C:	3			C:	2
A:	65	D:	2	C:	82	D:	8	D:	77	C:	13	A:	68	D:	5	B:	96	D:	0
		Total	33			Total:	16			Total	21			Total	30			Total:	2

MATCHING				
(2) N	(7) A	(9) H	(10) D	(1) K

Universidad Tecnológica Nacional

Licenciatura en Lengua Inglesa

Dissertation: Assessing Vocabulary: Does Test Format Influence Test Results?

Prof. Mariana Vanesa Castino - 2011

| C | | I | | C | | I | | C | | I | | C | | I | | C | | I | |
Answer	N° of students	Answer	N° of students	Answer	N° of students	Answer	N° of students	Answer	N° of students	Answer	N° of students	Answer	N° of students	Answer	N° of students	Answer	N° of students	Answer	N° of students
		B:	7			C:	2			B:	11			A:	13			A:	3
		C:	4			D:	7			D:	3			B:	10			B:	3
		D:	1			E:	1			E:	4			C:	3			D:	3
		E:	3			F:	1			G:	1			E:	1			E:	2
N:	77	G:	1	A:	75	G:	1	H:	34	I:	1	D:	43	F:	2	K:	60	F:	15
		H:	2			H:	4			j:	3			G:	12			H:	4
		I:	1			I:	2			K:	2			H:	1			J:	1
		L:	1			K:	4			M	30			I:	3			L:	2
		O	1			M	1			N:	1			J:	1			M	1
						N:	1			O	3			K:	3			O	1
						?:	1			?	5			L:	4			¿	3

	Total	21		Total:	23		Total:	64		M O Total	2 1 55		Total	38

ALTERNATE CHOICE

(I)				(B)				(D)				(H)				(C)			
C		*I*		*C*		*I*		*C*		*I*		*C*		*I*		*C*		*I*	
Answer	Nº of students	Answer	Nº of students	Answer	Nº of students	Answer	Nº of students	Answer	Nº of students	Answer	Nº of students	Answer	Nº of students	Answer	Nº of students	Answer	Nº of students	Answer	Nº of students
C	87		11	C	90	I	8	C	56	I	42	C	50	I	48	C	73	I	25

MULTIPLE-CHOICE CLOZE									
(6)				(7)				(8)	
C		I		C		I		C	
Answer	Nº of students	Answer	Nº of students	Answer	Nº of students	Answer	Nº of students	Answer	Nº of students
		A:	3			B:	0		
		B:	21			C:	2		
C:	63	D:	11	A:	95	D:	1	C:	45
		Total:	35			Total:	3		

(8)		(9)				(10)			
I		C		I		C		I	
Answer	Nº of students	Answer	Nº of students	Answer	Nº of students	Answer	Nº of students	Answer	Nº of students
A:	52			A:	20			A:	0
B:	1			B:	14			C:	12
D:	0	D:	41	C:	23	B:	78	D:	8
Total:	53			Total:	57			Total:	20

MATCHING				
(5) J	(3) G	(8) C	(6) M	(4) O

| C | | I | | C | | I | | C | | I | | C | | I | | C | | I | |
Answer	N° of students	Answer	N° of students	Answer	N° of students	Answer	N° of students	Answer	N° of students	Answer	N° of students	Answer	N° of students	Answer	N° of students	Answer	N° of students	Answer	N° of students
	15	C:	17			B:	5			B:	2			A:	1			E:	15
		D:	1			D:	6			D:	5			B:	28			F:	1
		H:	5			E:	3			F:	1			D:	11			H:	1
		K:	3			F:	2			G:	5			G:	4			J:	1
J:	70	F:	1	G:	62	H:	3	C:	50	H:	1	M	43	H:	1	O	79	?:	1
		L	1			I:	11			I:	6			I:	2				
						M	2			K:	1			K:	1				
						N:	1			L:	18			L:	8				
						¿:	4			M	2								
										N:	2								

		Total:	28			Total:	36			? Total:	5 48			Total:	55			Total:	19

ALTERNATE CHOICE									

(G)				(E)				(J)				(A)				(F)			
C		*I*		*C*		*I*		*C*		*I*		*C*		*I*		*C*		*I*	
Answer	*Nº of students*	*Answer*	*Nº of students*	*Answer*	*Nº of students*	*Answer*	*Nº of students*	*Answer*	*Nº of students*	*Answer*	*Nº of students*	*Answer*	*Nº of students*	*Answer*	*Nº of students*	*Answer*	*Nº of students*	*Answer*	*Nº of students*
C	80	I	18	C	88	I	10	C	50	I	48	C	59	I	39	C	93	I	5

Appendix 5: Data Matrix: Test results

Column key: **MCC** = Multiple-Choice Cloze, **MAT** = Matching, **ALT** = Alternate Choice. **S1–S10** = Score columns 1–10, **O1–O3** = Order columns 1–3.

Nª	Students	MCC S1	MCC S2	MCC S3	MCC S4	MCC S5	MCC S6	MCC S7	MCC S8	MCC S9	MCC S10	MCC O1	MCC O2	MCC O3	MAT S1	MAT S2	MAT S3	MAT S4	MAT S5	MAT S6	MAT S7	MAT S8	MAT S9	MAT S10	MAT O1	MAT O2	MAT O3	ALT S1	ALT S2	ALT S3	ALT S4	ALT S5	ALT S6	ALT S7	ALT S8	ALT S9	ALT S10	ALT O1	ALT O2	ALT O3
1	Patricio Rk				x									x					x						x								x							X	
2	Valentin P									x			x						x								x				x							X		
3	Florencia I				x							x						x								x					x									X
4	Nicole R					x								x			x								x									x				X		
5	Marìa Ailen G								x					x						x					x										x			X		
6	Andrès P									x		x								x						x									x					x
7	Santiago A								x			x										x				x									x					X
8	Malena M							x				x								x						x							x							X
9	Camila D									x				x					x						x						x							X		
10	Camila G									x			x									x					x								x			X		
11	Belen S							x					x						x								x						x					X		
12	Rosario D									x			x										x				x								x			x		
13	Cumelen D							x					x										x				x								x			X		
14	Pablo P							x						x					x						x								x					X		
15	Macarena S						x							x					x						x								x					x		
16	Facundo E						x					x										x				x									x					X
17	Franco F							x					x					x								x				x								X		
18	Natasha T						x							x								x		x									x					X		
19	Sofia D						x			x											x		x										x					x
20	Agustina B						x			x										x			x										x					X
21	Micaela P									x				x						x		x										x				X		
22	Sebastian B						x							x					x			x										x				X		
23	Martina U								x	x											x	x										x					X
24	Florencia R						x			x										x			x									x			X		
25	Macarena RR								x	x								x			x									x			X		
26	Mariana D			x				x				x							x			x							x			x		
27	Maria Jose D					x			x				x							x		x								x			X		
28	Tomàs B					x						x				x				x								x			X		
29	Florencia R					x		x								x			x							x				x
30	Victoria C				x			x							x			x							x				X
31	Luz A								x		x		x							x							x			X		
32	Juan Pablo T					x		x				x			x			x						x					X
33	Daniela A					x			x		x			x						x		X		
34	Eugenia S					x			x		x				x					x		X		
35	Andrea T			x		x			x				x					x		X
36	Malena C					x		x			x		x						x		X		
37	Daniela C			x		x			x		x			x					x		X		
38	Florencia ST					x		x			x		x					x		x		
39	Malena H				x		x		x		x					x		X		
40	Daniela O			x		x			x		x			x		X		
41	Fernando S	x		x		x		x		x			x			X
42	Agustina S				x		x		x		x			x		X		

Nª	Students	MULTIPLE-CHOICE CLOZE SCORE 1	2	3	4	5	6	7	8	9	10	ORDER 1	2	3	MATCHING SCORE 1	2	3	4	5	6	7	8	9	10	ORDER 1	2	3	ALTERNATE CHOICE SCORE 1	2	3	4	5	6	7	8	9	10	ORDER 1	2	3
43	Jeremias L			x								x				x											x						x								X
44	Julia M							x					x									x						x							x				X		
45	Denise M								x				x								x							x								x			X		
46	Juan M				x							x				x											x					x									X
47	Nicolàs A				x							x							x								x						x								X
48	Florencia P								x			x									x						x								x						X
49	Patricia B			x										x		x										x								x					X	
50	Sofia RM								x				x										x					x								x			X		
51	Rocio A				x									x							x					x								x					x	
52	Facundo L							x					x						x									x						x					X		
53	Laura P									x				x							x					x								x					x	
54	Hernan T			x									x			x												x						x					X		
55	Liliana T							x				x										x				x							x							X
56	Gabriel C						x					x											x			x								x						X
57	Pablo R							x				x										x				x								x						x
58	Lucas L				x									x		x										x								x					X	
59	Eugenia C								x				x											x			x								x			X		
60	Ana Sol F							x					x										x				x								x			X		
61	Ignacio C			x									x					x									x					x						x		
62	Bianca G						x					x								x						x						x								x
63	Francisco P							x				x				x										x						x								x
64	Nicolàs V			x								x								x						x						x					X		
65	Nahuel S				x									x		x										x								x			X		
66	Anahi R							x				x				x										x								x					X
67	Santiago N				x									x						x						x							x				X		
68	Daniela C						x							x							x					x								x			X		
69	Manuel B						x						x								x						x							x			X		
70	Joaquin NR					x							x			x											x				x						X		
71	Cristian I				x									x		x										x							x					x	
72	Juan V				x								x							x							x		x								X		
73	Julieta A				x								x							x							x						x				X		
74	Camila A					x						x								x						x								x					x
75	Manuel R							x				x								x						x							x						X
76	Julieta G O							x															x				x							x			X		
77	Constanza P					x						x								x							x						x				X		
78	Alfonso T								x					x						x					x							x					X	
79	Josefina V					x								x			x								x							x					X	
80	Josefina GM						x					x					x									x						x						X
81	Santiago A					x								x						x					x								x				X	
82	Santiago E					x								x					x					x								x				X	
83	Carola NF							x				x										x			x						x					X		
84	Matias C						x							x								x			x								x			X		
85	Guadalupe P			x								x				x											x						x						X

Universidad Tecnológica Nacional

Licenciatura en Lengua Inglesa

Dissertation: Assessing Vocabulary: Does Test Format Influence Test Results?

Prof. Mariana Vanesa Castino - 2011

Nª	Students	MULTIPLE-CHOICE CLOZE													MATCHING													ALTERNATE CHOICE												
		SCORE										ORDER			SCORE										ORDER			SCORE										ORDER		
		1	2	3	4	5	6	7	8	9	10	1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3
86	Florencia S			x								x							x								x						x								X
87	Sebastian M								x				x									x						x								x			x		
88	Angeles L								x					x				x								x									x					X	
89	Mercedes B									x			x									x						x									x		X		
90	Agostina I									x		x								x							x										x				X
91	Romina D							x					x							x								x								x			X		
92	Paloma I								x					x							x					x												x		X	
93	Leandro B							x				x								x							x									x					X
94	Lucia S										x			x							x					x								x						x	
95	Stefania A							x						x				x									x						x						X		
96	Agustina A						x					x						x									x											x			X
97	Gimena F								x					x									x				x								x				X		
98	Florencia B							x				x											x				x										x				X

Universidad Tecnológica Nacional

Licenciatura en Lengua Inglesa

Dissertation: Assessing Vocabulary: Does Test Format Influence Test Results?

Prof. Mariana Vanesa Castino - 2011

Appendix 6: Data Matrix: Interview results

Nª	Student	Question # 1				Question # 2			Question # 3						Question # 4					
		Task found more difficult				Task found easier			Helped choose correct option		Read sentence up to blank & read 4 options	Read whole sentence before the 4 options	Read whole text before 4 options	Read whole paragr. Before 4 options	Thought of a context to include the word	Did it at random	1st matched the words known as a mean of diserching	Had to repeat words several times	Had to resort to translation	Looked at syntactic function of words
		None of the 3	Alternate & Matching	Alternate Choice	Matching	Multiple-choice cloze	Alternate choice	Matching	Meaning words in context	Syntactic function in particular cases										
1	Macarena RR	X				X			x	X			X		X					
2	Franco F			X		X			x	X		X				X				
3	Valentin P			X		X			x	X	X					X				
4	Nicole R				X	X			x	X	X					X				
5	Eugenia S			X				x	x	X	X				X					
6	Tomas B				X	X			x	X		X				X				
7	Mariana D				X		X		x	X			X				X			
8	Luz A				X	X			x	X	X						X			
9	Camila D		x			X			x	X	X				X					
10	Pablo P				X	X			x	X	X				X					
11	Andrès P				X	X			x	X	X				X					
12	Macarena S			X					x	X	X									X
13	Florencia ST				X	X			x	X			X			X				
14	Andrea T				X		X		x	X			X			X				
15	Rocio A				X	X			x	X			X		X					
16	Patricia B				X		X		x	X			X			x				
17	Facundo L			X		X			x	X	X						X			
18	Hernan T				X		X		x	X				X	X					
19	Lucas L				X	X			x	X		X								X
20	Joaquin NR				X	X			x	X	X				X					
21	Daniela C				X	X			x	X		X					X			
22	Bianca G			X		X			x	X			X					X		
23	Anahi R				X	X			x	X			X					X		
24	Cristian I				X		X		x	X						X		x		
25	Constanza P				X		X		x	X	X				X					
26	Camila A				X	X			x	X	X									X
27	Liliana T			X		X			x	X	X				X					
28	Manuel R					X			x	X					X	x				
29	Guadalupe P					X			x	X			X		X					
30	Angeles L				X	X			x	X	X							X		
31	Agostina I				X	X			x	X		X			X					
32	Lucia S				X	X			x	X		x			X					
33	Agustina A				X	X			x	X			X		x					

Nª	Student	Question # 5			Question # 6 More familiar with			Question # 7 Less familiar with				Question # 8 Practised more			Question # 9 Practised less		
		Thought they knew meaning without guessing	Had to think of an example	Used guessing strategies	Multiple-choice cloze	Alternate choice	Matching	Multiple-choice cloze	Alternate choice	Matching	None	Multiple-choice cloze	Alternate choice	Matching	Multiple-choice cloze	Alternate choice	Matching
1	Macarena RR	X			X						X	X				X	
2	Franco F			X	X				X			X				X	
3	Valentin P			X	X				X			X				X	
4	Nicole R	X			X				X			X				X	
5	Eugenia S		X		X				X			X				X	
6	Tomas B	X			X				X	x		X					x
7	Mariana D			X	X				X			X				X	
8	Luz A		X		X				X			X				X	
9	Camila D		X		X				X			X				X	
10	Pablo P		X		X				X	X		X				X	
11	Andrès P	X			X					X		X					x
12	Macarena S			X	X				X			X					X
13	Florencia ST			X	X				X			X					X
14	Andrea T		X		X					x		X					X
15	Rocio A	X			x				X			X				X	
16	Patricia B	X		X		X				X		X					X
17	Facundo L				X					X		X					X
18	Hernan T	X			X					X		X					X
19	Lucas L	X			X				X			X				X	
20	Joaquin NR	X			X				X			X				X	
21	Daniela C	X			X				X			X				X	
22	Bianca G			x	X				X			X				X	
23	Anahi R	X			X				X			X				X	
24	Cristian I	X			X	x				X		X					X
25	Constanza P	X				X				X		X					X
26	Camila A	X			x	X				X		X					X
27	Liliana T	X			X					X		X					X
28	Manuel R	X			X					X		X					X
29	Guadalupe P	X				x				X		X					X
30	Angeles L	X			X				X			X				X	
31	Agostina I		x		X					X		X					X
32	Lucia S	X			X				X			X				X	
33	Agustina A	X			X					X		X					X

Appendix 7: Data Matrix: Interview responses

N°	Student	Question # 1	Question # 2	Question # 3	Question # 4	Question # 5
1	Macarena RR	None of the three looked really difficult	The cloze was easy because the context helped.	I centred on the text first. I always read the text first	I had to think of a context	I had to think more than in the cloze but it was not difficult
2	Franco F	It was confussing. Had to read it (C/I) several times for not finding the words in a context.	The cloze was easier because the context helped.	The text is what helped me more. Also the meaning and the syntax in some cases. I read the sentence and the 4 options.	It was difficult. I needed to know more vocabulary. I think I did it quite at random.	It was confussing "daba muchas vueltas". It didn´t have a context.
3	Valentin P	The words, the meaning, the definitions were confussing in the Altenate choice.	The general context where to place the words helped. Placed the words as it sounded in the cloze.	First I see part of the sentence and then I pay attention to the options.	It was done at random matching the synonyms as it sounded to me.	This one confused me more I did it at random.
4	Nicole R	It was difficult for having to match synonyms. It was difficult.	Found the cloze easier for finding meaning in the context and for being used to doing it.	The context helped. If you know what the text talks about you should know which word fits.	I did it at random. It was quite difficult.	I did it as I thought it was correct. I understood.
5	Eugenia S	The alternate choice was difficult for not having a sentence in context as in the cloze.	Found the matching easier for the fact of finding the synonyms in the other column	In some cases collocations helped me discard. The fact of having 4 options confused me. I read up to the blank	I had to thank of a sentence.	I resorted to sentences. I always think of sentences.
6	Tomas B	It was difficult for having been the first.	The cloze was easier for having done it at last.	I repeat the word in a sentence to see how it sounds and if it fits.	I did it quite quickly to hand in. Seeing similar words confused me.	Remembering having heard the word in that use marks the difference. Eg, "to cope with, to manage".
7	Mariana D	It was confussing. All the words looked similar.	The fact of having to say if it was right or wrong with definitions. It was the one with fewer chances to make mistakes (Alternate choice)	I read it. I remembered some words. I tried to see which of them were collocations. I paid attention to both: text and 4 options.	I started from the words that i knew and then I passed on to discarding.	Because of the fact of having to say "right or wrong". I had less chances to make mistakes.
8	Luz A	It was difficult for being the first.	The cloze was not easier for having being the third but for the context.	The context helped me more because I knew the words context and syntax helped me. In some cases collocations helped. When I came to the blank I read the options and knew which word to fit in matching.	First, I matched the words I kew and then by reading I matched the rest. This process was uncouncious.	I thought of examples where to use that word.
9	Camila D	The matching was as difficult as the Alterntate. Had to think of a context.	In the cloze, because of the context and being used to doing it found it easier.	From the context and the syntax I realised which word could fit. I knew that "recognize" was followed by "to".	I tried to use a word form the second column in a context and then replace this word by a word in the first column in the same sentence.	I made it more difficult. I looked for something that said exactly the same.
10	Pablo P	There wasn´t any example of a context.	Found the cloze easier for the context.	As I know most of the words I know how to use each in different contexts. The syntax also helped me decide in some cases.	I had to think of a context myself where to place the words.	I said "I can use this word with that and that in a sentence.
11	Andrès P	It is easier to find words in a context.	Found the cloze easier for the context.	One decides which words to fill in with as one reads the sentence or text and the options.	First I tried to imagine the word in a context and that helped me get to know the meaning and the synonyms of that word.	I knew the meaning. It has to do with vocabulary already learnt and you can learn it with this type of exercises.

N°	Student	Question # 1	Question # 2	Question # 3	Question # 4	Question # 5
12	Macarena S	The definitions were contradictory in the Alternate choice	Found the cloze easier because it had a context.	The context shows what the sentence aims at so that you can continueit with a word.	I had to resort to translate into Spanish.	I did it quite at random.
13	Florencia ST	The meaning of words can be understood from the context where then can be found. Meaning depend on context.	I´m more used to doing the cloze. The context helped in some way (cloze)	First I read the whole text to see what it was about and then I read the options.	I don´t think it was the most complicated for being the fires but then when I did the others I realized that I had heard some words.	It was more a question of luck
14	Andrea T	It was confussing for having to match synonyms. One has to know a lot of vocabulary in order to be able to do this.	I could get the meaning easier. The words were not difficult (Alternate choice).	The meaning can be deduced from the context. Te syntax helps with certain patterns. I concentrated on the text first.	It was done completely at random.	The difference with the cloze is that there you get the meaning of the word from the text.
15	Rocio A	It was too exact. Difficult for not having the words in context.	You can get the answer in he cloze because of the context.	The text guides you. I first read the text and then the options. I first think of the word and then I look at the options.	I had to resort to a context. you have to do this type of exercises with examples.	It had to do with previous knowledge.
16	Patricia B	Difficult for having the words out of the context.	The meaning made me remember what the words meant (Alternate choice). The fact of being the second made it easier.	I first read the whole text and then the options. They were confussing because they were similar.	I did it at random	A text or a sentence always helps. The sentence helped remember another sentence.
17	Facundo L	It was the most difficult for having been the first.	I always do well in this type of exercise. The context helped me (cloze)	I read the text and when I had the space I read the options	First I discarded the words I knew. It was difficult for not having a context or a definition	In some I had to guess if it was correct or incorrect.
18	Hernan T	It was confussing with words difficult to understand.	The definition helped me quite a lot (Alternative choice).	First I read the whole paragraph. I first looked at the context and repeated the word in the sentence.	I thought if the word had the same meaning and tried to find an example.	I did it as I thought it was correct.
19	Lucas L	Difficult for finding more options in one column	Because of having seen the previous ones. It (cloze) was the third and we practised it in use of English (cloze).	I paid attention to the text and the 4 options. I first read the sentence and then the options.	I had to resort to Spanish.	I did it by resorting.
20	Joaquin NR	Confussing for not finding similar between some words.	For being used to doing this type of exercises. The text helped me (cloze).	The text helped me more than I thought. The 4 options confused me.	I tried to see what I could do with the words.	It wasn´t at ramdom. I tried to understand.
21	Daniela C	Difficult for not having the words in a sentence.	For being within a text it´s normally the one that is usually easier (cloze).	The word surrounding the blank helped when I read the sentence, the options and the context.	First I matched the words that I knew and the the ones that sounded familiar to me.	I remembered the words from the classes (eg "to manage") and the definition helped.
22	Bianca G	The definition were confussing	I did it faster because of the context (cloze).	I see the text first and then I fill in what sounds better.	I read the words and I repeated them to see it. They sounded familiar to me.	I found it very difficult. It was too much grammar for me.
23	Anahi R	Several words looked similar.	For having a context. For getting the answers from the context (cloze)	First I read the whole text. There was a case in which the syntax helped "with".	I tried to think of a synonym repeating the words.	I know the meaning of almost all the points.
24	Cristian I	Did for not having a sentence or context where to place the word.	It´s easier to understand a definition (Alternate choice).	I read the whole paragraph in most of the cases. I founded on the text first and then the options.	I didn´t think of a context. I tried to understand the meaning of each words.	With the context I could understand it better.
25	Constanza P	Some words looked very similar.	It was easier for being a definition (Alternate choice).	I start reading and when I get to the blank I read the options. I never read the whole text, in fact I didn´t understand.	I had to think of a sentence.	I didn´t have to think of a sentence. It was enough with the definition.

N°	Student	Question # 1	Question # 2	Question # 3	Question # 4	Question # 5
26	Camila A	Did not find a context where to place the word.	Maybe I did better en the cloze for having being the third (cloze).	The context and syntax helped me.	I started doing the exercise by looking at the syntactic function. I knew I couldn´t match "quantities" with a verb.	I focused on the whole sentence.
27	Liliana T	The sentences in the Alternate choice were confussing for being out of context.	The factor of being more used to doing it really have influence. One has another context in the cloze.	I usually read the text till where I have to fill in the blank. The syntax helps when in doubt with a word.	This is the first time I´ve done this type of exercise. The lack of context complicated things.	I think I knew the meaning.
28	Manuel R	There were words difficult to understand out of context.	Because it was the one to choose (Alternate choice) with sentences.	I read sentence by sentence and maybe finished the paragraph. I think of the sense of the sentence.	I placed the words in sentences.	With these sentences it was enough.
29	Guadalupe P	It was difficult without a context.	Because in this exercise (Alternate choice) you have the words and the meaning in sentences.	I read the text twice before filling the blanks. The first time I placed what I thought was right.	I thought of an example with the word,	It is better when the meaning is in sentences.
30	Angeles L	It is necessary to have a context	Because when I read the sentence I can see what sounds better (cloze).	I always see if there´s a key word before or after the blank and I also look at the meaning of the sentence. The syntax helped a lot.	I looked for the synonym of the ones I knew and I saw it in the other column there was a word that was a similar to the one I didn´t know from the first column.	I paid attention to how the sentence sounded to me.
31	Agostina I	For not being used to doing these exercises without a context.	Even though I was not sure about some words I got them from what came next in the sentence (cloze).	I read the sentence and then the options.	I tried to look for something where I could replace the word.	I thought of a specific meaning in sentences.
32	Lucia S	For having to place the synonyms out of context.	Because I could put which word fit before or after the blank (cloze).	I focused on the sentence first and then the options. In some cases I could say which word could fit before reading the options especially because of collocations.	I tried to make a sentence and then with the other word I saw if the meaning was the same.	I looked at the meaning of both words and saw if it matched with the meaning of the whole sentence.
33	Agustina A	It was difficult to associate the words without a context.	Because of having a context (cloze).	I read the text then the complete sentence and filled in the blank.	For some words I thought of an example.	I knew the meaning.

N°	Student	Question # 6	Question # 7	Question # 7	Question # 9
1	Macarena RR				
2	Franco F				
3	Valentin P				
4	Nicole R				
5	Eugenia S	I do the matching in reading tasks. Match words with definitions.			I´ve never done the alternate-choice.
6	Tomas B		We don´t do the Alternate-choice or marching here. Then appear in books to practice vocabulary.		
7	Mariana D	I do the cloze more often.			
8	Luz A				This is the firs time I´ve done it (Alternate choice) here.
9	Camila D				
10	Pablo P				I saw the matching in the book, not in the mock.

Licenciatura en Lengua Inglesa

Dissertation: Assessing Vocabulary: Does Test Format Influence Test Results?

Prof. Mariana Vanesa Castino - 2011

N°	Student	Question # 6	Question # 7	Question # 7	Question # 9
11	Andrès P				We´ve never done this type of exercise (Alternate choice)
12	Macarena S				
13	Florencia ST			I do the cloze in mocks.	
14	Andrea T				
15	Rocio A				
16	Patricia B				
17	Facundo L	We practised it a lot because of the FCE			
18	Hernan T				
19	Lucas L				

Universidad Tecnológica Nacional

Licenciatura en Lengua Inglesa

Dissertation: Assessing Vocabulary: Does Test Format Influence Test Results?

Prof. Mariana Vanesa Castino - 2011